

10/10/00
Customer No. 20350
TOWNSEND and TOWNSEND and CREW LLP
Embarcadero Center, 8th Floor
San Francisco, California 94111-3834
76-0200

10-12-00

A
JC921 U.S.P.T.O.
09/688672
10/10/00

**BOX PATENT APPLICATION
ASSISTANT COMMISSIONER FOR PATENTS
Washington, D.C. 20231**

Sir:

Transmitted herewith for filing under 37 CFR 1.53(b) is the
[X] patent application of
[] continuation patent application of
[] divisional patent application of
[] continuation-in-part patent application of

Attorney Docket No. 014058-009041US

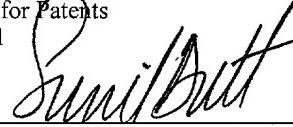
Client Ref No. CX-99-0045 and CX-99-0061

"Express Mail" Label No. EL394 879 915US

Date of Deposit: October 10, 2000

I hereby certify that this is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above, addressed to:

**BOX PATENT APPLICATION
Assistant Commissioner for Patents
Washington, D.C. 20231**

By: 

Inventors/Applicants Identifier: Yasir Skeiky, Steven Reed, Raymond L. Houghton, Patricia D. McNeill, Davin C. Dillon, and Michael J. Lodes

For: FUSION PROTEINS OF MYCOBACTERIUM TUBERCULOSIS

[X] This application claims priority from each of the following Application Nos./filing dates:
60/158,338, filed October 7, 1999; and 60/158,425, filed October 7, 1999,
the disclosures of which are incorporated by reference.

Enclosed are:

- [X] 80 pages of specification
[X] 14 pages of claims
[X] 1 page of Abstract
[X] 1 Title Page
[X] 38 sheets of [] formal [X] informal drawings.
[X] 31 pages of informal sequence listing.

Two assignments of the invention from the priority applications to Corixa Corporation

An [] signed [X] unsigned Declaration.

A Power of Attorney by Assignee with Certificate Under 37 CFR Section 3.73(b).

A certified copy of a _____ application.

Information Disclosure Statement under 37 CFR 1.97.

A petition to extend time to respond in the parent application.

Notification of change of [] power of attorney [] correspondence address filed in prior application.

(Col. 1)	(Col. 2)
FOR:	NO. FILED
BASIC FEE	NO. EXTRA
TOTAL CLAIMS	115 - 20 = *95
INDEP. CLAIMS	24 - 3 = *21
[] MULTIPLE DEPENDENT CLAIM PRESENTED	

* If the difference in Col. 1 is less than 0, enter "0" in Col. 2.

Please charge Deposit Account No. 20-1430 as follows:

- [X] Filing fee \$ \$4,100.00
[X] Any additional fees associated with this paper or during the pendency of this application.

2 extra copies of this sheet are enclosed.

Telephone:
(415) 576-0200

Facsimile:
(415) 576-0300

SMALL ENTITY	
RATE	FEE
	\$355.00
x \$9.00 =	
x \$40.00 =	
+ \$135.00 =	
TOTAL	

OTHER THAN SMALL ENTITY	
RATE	FEE
	\$710.00
x \$18.00 =	\$1,710.00
x \$80.00 =	\$1,680.00
+ \$270.00 =	
TOTAL	\$4,100.00

Respectfully submitted,
TOWNSEND and TOWNSEND and CREW LLP


Annette S. Parent

Reg No. 42,058
Attorneys for Applicants

Attorney Docket No.: 014058-009041US
Client Ref. Nos. CX-99-0045 and CX-99-0061

PATENT APPLICATION

FUSION PROTEINS OF MYCOBACTERIUM TUBERCULOSIS

Inventor(s):

Yasir Skeiky, a citizen of Canada and Lebanon, residing at,
8327 25th Avenue, NW
Seattle, WA 98107

Steven Reed, a citizen of United States, residing at,
2843 122nd Place, NE
Bellevue, WA 98005

Raymond L. Houghton, a citizen of United States, residing at,
2636 242nd Place, SE
Bothell, WA 98021

Patricia D. McNeill, a citizen of United States, residing at
1421 S. 248th Street
Des Moines, WA 98198

Davin C. Dillon, a citizen of United States, residing at
18112 NW Montreux Drive
Issaquah, Washington 98027

Michael J. Lodes, a citizen of United States, residing at,
9223 36th Avenue, SW
Seattle, WA 98126

Assignee:

Corixa Corporation
1124 Columbia Street, Suite 200
Seattle, WA 98104

Entity: Large

TOWNSEND and TOWNSEND and CREW LLP
Two Embarcadero Center, 8th Floor
San Francisco, California 94111-3834
Tel: 415-576-0200

FUSION PROTEINS OF MYCOBACTERIUM TUBERCULOSIS

CROSS-REFERENCES TO RELATED APPLICATIONS

The present application claims priority to U.S. patent application No.

- 5 60/158,338, filed October 7, 1999, and U.S. application No. 60/158,425, filed October 7, 1999, herein each incorporated by reference in its entirety.

This application is also related to U.S. patent application No. 09/056,556, filed April 7, 1998; U.S. patent application No. 09/223,040, filed December 30, 1998; U.S. patent application No. 09/287,849, filed April 7, 1999; and published PCT application No.

- 10 WO99/51748, filed April 7, 1999 (PCT/US99/07717), herein each incorporated by reference in its entirety.

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

- 15 Not applicable.

BACKGROUND OF THE INVENTION

Tuberculosis is a chronic infectious disease caused by infection with *M. tuberculosis* and other Mycobacterium species. It is a major disease in developing countries, 20 as well as an increasing problem in developed areas of the world, with about 8 million new cases and 3 million deaths each year. Although the infection may be asymptomatic for a considerable period of time, the disease is most commonly manifested as an acute inflammation of the lungs, resulting in fever and a nonproductive cough. If untreated, serious complications and death typically result.

25 Although tuberculosis can generally be controlled using extended antibiotic therapy, such treatment is not sufficient to prevent the spread of the disease. Infected individuals may be asymptomatic, but contagious, for some time. In addition, although compliance with the treatment regimen is critical, patient behavior is difficult to monitor. Some patients do not complete the course of treatment, which can lead to ineffective 30 treatment and the development of drug resistance.

In order to control the spread of tuberculosis, effective vaccination and accurate early diagnosis of the disease are of utmost importance. Currently, vaccination with live bacteria is the most efficient method for inducing protective immunity. The most

DOCUMENT NUMBER: 10000000000000000000000000000000

common mycobacterium employed for this purpose is Bacillus Calmette-Guerin (BCG), an avirulent strain of *M. bovis*. However, the safety and efficacy of BCG is a source of controversy and some countries, such as the United States, do not vaccinate the general public with this agent.

5 Diagnosis of tuberculosis is commonly achieved using a skin test, which involves intradermal exposure to tuberculin PPD (protein-purified derivative). Antigen-specific T cell responses result in measurable induration at the injection site by 48-72 hours after injection, which indicates exposure to mycobacterial antigens. Sensitivity and specificity have, however, been a problem with this test, and individuals vaccinated with
10 BCG cannot be distinguished from infected individuals.

While macrophages have been shown to act as the principal effectors of Mycobacterium immunity, T cells are the predominant inducers of such immunity. The essential role of T cells in protection against Mycobacterium infection is illustrated by the frequent occurrence of Mycobacterium infection in AIDS patients, due to the depletion of
15 CD4⁺ T cells associated with human immunodeficiency virus (HIV) infection.
Mycobacterium-reactive CD4⁺ T cells have been shown to be potent producers of γ -interferon (IFN- γ), which, in turn, has been shown to trigger the anti-mycobacterial effects of macrophages in mice. While the role of IFN- γ in humans is less clear, studies have shown that 1,25-dihydroxy-vitamin D3, either alone or in combination with IFN- γ or tumor necrosis
20 factor-alpha, activates human macrophages to inhibit *M. tuberculosis* infection. Furthermore, it is known that IFN- γ stimulates human macrophages to make 1,25-dihydroxy-vitamin D3. Similarly, interleukin-12 (IL-12) has been shown to play a role in stimulating resistance to *M. tuberculosis* infection. For a review of the immunology of *M. tuberculosis* infection, see Chan & Kaufmann, *Tuberculosis: Pathogenesis, Protection and Control* (Bloom ed., 1994),
25 and Harrison's *Principles of Internal Medicine*, volume 1, pp. 1004-1014 and 1019-1023 (14th ed., Fauci et al., eds., 1998).

Accordingly, there is a need for improved diagnostic reagents, and improved methods for diagnosis, preventing and treating tuberculosis.

30 SUMMARY OF THE INVENTION

The present invention provides pharmaceutical compositions comprising at least two heterologous antigens, fusion proteins comprising the antigens, and nucleic acids encoding the antigens, where the antigens are from a *Mycobacterium* species from the

tuberculosis complex and other *Mycobacterium* species that cause opportunistic infections in immune compromised patients. The present invention also relates to methods of using the polypeptides and polynucleotides in the diagnosis, treatment and prevention of *Mycobacterium* infection.

5 The present invention is based, in part, on the inventors' discovery that fusion polynucleotides, fusion polypeptides, or compositions that contain at least two heterologous *M. tuberculosis* coding sequences or antigens are highly antigenic and upon administration to a patient increase the sensitivity of tuberculosis sera. In addition, the compositions, fusion polypeptides and polynucleotides are useful as diagnostic tools in patients that may have been
10 infected with *Mycobacterium*.

In one aspect, the compositions, fusion polypeptides, and nucleic acids of the invention are used in *in vitro* and *in vivo* assays for detecting humoral antibodies or cell-mediated immunity against *M. tuberculosis* for diagnosis of infection or monitoring of disease progression. For example, the polypeptides may be used as an *in vivo* diagnostic
15 agent in the form of an intradermal skin test. The polypeptides may also be used in *in vitro* tests such as an ELISA with patient serum. Alternatively, the nucleic acids, the compositions, and the fusion polypeptides may be used to raise anti-*M. tuberculosis* antibodies in a non-human animal. The antibodies can be used to detect the target antigens *in vivo* and *in vitro*.

20 In another aspect, the compositions, fusion polypeptides and nucleic acids may be used as immunogens to generate or elicit a protective immune response in a patient. The isolated or purified polynucleotides are used to produce recombinant fusion polypeptide antigens *in vitro*, which are then administered as a vaccine. Alternatively, the polynucleotides may be administered directly into a subject as DNA vaccines to cause
25 antigen expression in the subject, and the subsequent induction of an anti-*M. tuberculosis* immune response. Thus, the isolated or purified *M. tuberculosis* polypeptides and nucleic acids of the invention may be formulated as pharmaceutical compositions for administration to a subject in the prevention and/or treatment of *M. tuberculosis* infection. The immunogenicity of the fusion proteins or antigens may be enhanced by the inclusion of an
30 adjuvant, as well as additional fusion polypeptides, from *Mycobacterium* or other organisms, such as bacterial, viral, mammalian polypeptides. Additional polypeptides may also be included in the compositions, either linked or unlinked to the fusion polypeptide or compositions.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the nucleic acid sequence of a vector encoding TbF14 (SEQ ID NO:89). Nucleotides 5096 to 8594 encode TbF14 (SEQ ID NO:51). Nucleotides 5072 to 5095 encode the eight amino acid His tag (SEQ ID NO:90); nucleotides 5096 to 7315 encode 5 the MTb81 antigen (SEQ ID NO:1); and nucleotides 7316 to 8594 encode the Mo2 antigen (SEQ ID NO:3).

Figure 2 shows the nucleic acid sequence of a vector encoding TbF15 (SEQ ID NO:91). Nucleotides 5096 to 8023 encode the TbF15 fusion protein (SEQ ID NO:53). Nucleotides 5072 to 5095 encode the eight amino acid His tag region (SEQ ID NO:90); 10 nucleotides 5096 to 5293 encode the Ra3 antigen (SEQ ID NO:5); nucleotides 5294 to 6346 encode the 38 kD antigen (SEQ ID NO:7); nucleotides 6347 to 6643 encode the 38-1 antigen (SEQ ID NO:9); and nucleotides 6644 to 8023 encode the FL TbH4 antigen (SEQ ID NO:11).

Figure 3 shows the amino acid sequence of TbF14 (SEQ ID NO:52), including 15 the eight amino acid His tag at the N- terminus.

Figure 4 shows the amino acid sequence of TbF15 (SEQ ID NO:54), including the eight amino acid His tag at the N-terminus.

Figure 5 shows ELISA results using fusion proteins of the invention.

Figure 6 shows the nucleic acid and the predicted amino acid sequences of the 20 entire open reading frame of HTCC#1 FL (SEQ ID NO:13 and 14, respectively).

Figure 7 shows the nucleic acid and predicted amino acid sequences of three fragments of HTCC#1. (a) and (b) show the sequences of two overlapping fragments: an amino terminal half fragment (residues 1 to 223), comprising the first trans-membrane domain (a) and a carboxy terminal half fragment (residues 184 to 392), comprising the last 25 two trans-membrane domains (b); (c) shows a truncated amino-terminal half fragment (residues 1 to 128) devoid of the trans-membrane domain.

Figure 8 shows the nucleic acid and predicted amino acid sequences of a TbRa12-HTCC#1 fusion protein (SEQ ID NO:63 and 64, respectively).

Figure 9a shows the nucleic acid and predicted amino acid sequences of a 30 recombinant HTCC#1 lacking the first trans-membrane domain (deleted of the amino acid residues 150 to 160). Figure 9b shows the nucleic acid and predicted amino acid sequences of 30 overlapping peptides of HTCC#1 used for the T-cell epitope mapping. Figure 9c illustrates the results of the T-cell epitope mapping of HTCC#1. Figure 9d shows the nucleic

DOCUMENT-20200606

acid and predicted amino acid sequences of a deletion construct of HTCC#1 lacking all the trans-membrane domains (deletion of amino acid residues 101 to 203).

Figure 10 shows the nucleic acid and predicted amino acid sequences of the fusion protein HTCC#1(184-392)-TbH9-HTCC#1(1-129) (SEQ ID NO:57 and 58, 5 respectively).

Figure 11 shows the nucleic acid and predicted amino acid sequences of the fusion protein HTCC#1(1-149)-TbH9-HTCC#1(161-392) (SEQ ID NO:59 and 60, respectively).

Figure 12 shows the nucleic acid and predicted amino acid sequences of the 10 fusion protein HTCC#1(184-392)-TbH9-HTCC#1(1-200) (SEQ ID NO:61 and 62, respectively).

Figure 13 shows the nucleotide sequence of *Mycobacterium tuberculosis* antigen MTb59 (SEQ ID NO:49).

15 Figure 14 shows the amino acid sequence of *Mycobacterium tuberculosis* antigen MTb59 (SEQ ID NO:50).

Figure 15 shows the nucleotide sequence of *Mycobacterium tuberculosis* antigen MTb82 (SEQ ID NO:47).

Figure 16 shows the amino acid sequence of *Mycobacterium tuberculosis* antigen MTb82 (SEQ ID NO:48).

20 Figure 17 shows the amino acid sequence of *Mycobacterium tuberculosis* the secreted form of antigen DPPD (SEQ ID NO:44).

DESCRIPTION OF SEQUENCES

SEQ ID NO:1 is the nucleic acid sequence encoding the Mtb81 antigen.

25 SEQ ID NO:2 is the amino acid sequence of the Mtb81 antigen.

SEQ ID NO:3 is the nucleic acid sequence encoding the Mo2 antigen.

SEQ ID NO:4 is the amino acid sequence of the Mo2 antigen.

SEQ ID NO:5 is the nucleic acid sequence encoding the TbRa3 antigen.

SEQ ID NO:6 is the amino acid sequence of the TbRa3 antigen.

30 SEQ ID NO:7 is the nucleic acid sequence encoding the 38kD antigen.

SEQ ID NO:8 is the amino acid sequence of the 38kD antigen.

SEQ ID NO:9 is the nucleic acid sequence encoding the Tb38-1 antigen.

SEQ ID NO:10 is the amino acid sequence of the Tb38-1 antigen.

SEQ ID NO:11 is the nucleic acid sequence encoding the full-length (FL) TbH4 antigen.

SEQ ID NO:12 is the amino acid sequence of the FL TbH4 antigen.

5 SEQ ID NO:13 is the nucleic acid sequence encoding the HTCC#1 (Mtb40) antigen.

SEQ ID NO:14 is the amino acid sequence of the HTCC#1 antigen.

SEQ ID NO:15 is the nucleic acid sequence of an amino terminal half fragment (residues 1 to 223) of HTCC#1, comprising the first trans-membrane domain.

10 SEQ ID NO:16 is the predicted amino acid sequence of an amino terminal half fragment (residues 1 to 223) of HTCC#1.

SEQ ID NO:17 is the nucleic acid sequence of a carboxy terminal half fragment (residues 184 to 392) of HTCC#1, comprising the last two trans-membrane domains.

15 SEQ ID NO:18 is the predicted amino acid sequence of a carboxy terminal half fragment (residues 184 to 392) of HTCC#1.

SEQ ID NO:19 is the nucleic acid sequence of a truncated amino-terminal half fragment (residues 1 to 128) of HTCC#1 devoid of the trans-membrane domain.

20 SEQ ID NO:20 is the predicted amino acid sequence of a truncated amino-terminal half fragment (residues 1 to 128) of HTCC#1.

SEQ ID NO:21 is the nucleic acid sequence of a recombinant HTCC#1 lacking the first trans-membrane domain (deleted of the amino acid residues 150 to 160).

25 SEQ ID NO:22 is the predicted amino acid sequence of a recombinant HTCC#1 lacking the first trans-membrane domain (deleted of the amino acid residues 150 to 160).

SEQ ID NO:23 is the nucleic acid sequence of a deletion construct of HTCC#1 lacking all the trans-membrane domains (deletion of amino acid residues 101 to 203).

30 SEQ ID NO:24 is the predicted amino acid sequence of a deletion construct of HTCC#1 lacking all the trans-membrane domains (deletion of amino acid residues 101 to 203).

SEQ ID NO:25 is the nucleic acid sequence encoding the TbH9 (Mtb39A) antigen.

SEQ ID NO:26 is the amino acid sequence of the TbH9 antigen.

SEQ ID NO:27 is the nucleic acid sequence encoding the TbRa12 antigen.

- SEQ ID NO:28 is the amino acid sequence of the TbRa12 antigen.
- SEQ ID NO:29 is the nucleic acid sequence encoding the TbRa35 (Mtb32A) antigen.
- 5 SEQ ID NO:30 is the amino acid sequence of the TbRa35 antigen.
- SEQ ID NO:31 is the nucleic acid sequence encoding the MTCC#2 (Mtb41) antigen.
- SEQ ID NO:32 is the amino acid sequence of the MTCC#2 antigen.
- 10 SEQ ID NO:33 is the nucleic acid sequence encoding the MTI (Mtb9.9A) antigen.
- SEQ ID NO:34 is the amino acid sequence of the MTI antigen.
- 15 SEQ ID NO:35 is the nucleic acid sequence encoding the MSL (Mtb9.8) antigen.
- SEQ ID NO:36 is the amino acid sequence of the MSL antigen.
- 15 SEQ ID NO:37 is the nucleic acid sequence encoding the DPV (Mtb8.4) antigen.
- SEQ ID NO:38 is the amino acid sequence of the DPV antigen.
- SEQ ID NO:39 is the nucleic acid sequence encoding the DPEP antigen.
- SEQ ID NO:40 is the amino acid sequence of the DPEP antigen.
- 20 SEQ ID NO:41 is the nucleic acid sequence encoding the Erd14 (Mtb16) antigen.
- SEQ ID NO:42 is the amino acid sequence of the Erd14 antigen.
- SEQ ID NO:43 is the nucleic acid sequence encoding the DPPD antigen.
- SEQ ID NO:44 is the amino acid sequence of the DPPD antigen.
- 25 SEQ ID NO:45 is the nucleic acid sequence encoding the ESAT-6 antigen.
- SEQ ID NO:46 is the amino acid sequence of the ESAT-6 antigen.
- SEQ ID NO:47 is the nucleic acid sequence encoding the Mtb82 (Mtb867) antigen.
- 30 SEQ ID NO:48 is the amino acid sequence of the Mtb82 antigen.
- SEQ ID NO:49 is the nucleic acid sequence encoding the Mtb59 (Mtb403) antigen.
- SEQ ID NO:50 is the amino acid sequence of the Mtb59 antigen.
- SEQ ID NO:51 is the nucleic acid sequence encoding the TbF14 fusion protein.
- SEQ ID NO:52 is the amino acid sequence of the TbF14 fusion protein.

SEQ ID NO:53 is the nucleic acid sequence encoding the TbF15 fusion protein.

SEQ ID NO:54 is the amino acid sequence of the TbF15 fusion protein.

SEQ ID NO:55 is the nucleic acid sequence of the fusion protein

5 HTCC#1(FL)-TbH9(FL).

SEQ ID NO:56 is the amino acid sequence of the fusion protein HTCC#1(FL)-TbH9(FL).

SEQ ID NO:57 is the nucleic acid sequence of the fusion protein HTCC#1(184-392)-TbH9-HTCC#1(1-129).

10 SEQ ID NO:58 is the predicted amino acid of the fusion protein HTCC#1(184-392)-TbH9-HTCC#1(1-129).

SEQ ID NO:59 is the nucleic acid sequence of the fusion protein HTCC#1(1-149)-TbH9-HTCC#1(161-392).

15 SEQ ID NO:60 is the predicted amino acid sequence of the fusion protein HTCC#1(1-149)-TbH9-HTCC#1(161-392).

SEQ ID NO:61 is the nucleic acid sequence of the fusion protein HTCC#1(184-392)-TbH9-HTCC#1(1-200).

SEQ ID NO:62 is the predicted amino acid sequence of the fusion protein HTCC#1(184-392)-TbH9-HTCC#1(1-200).

20 SEQ ID NO:63 is the nucleic acid sequence of the TbRa12-HTCC#1 fusion protein.

SEQ ID NO:64 is the predicted amino acid sequence of the TbRa12-HTCC#1 fusion protein.

25 SEQ ID NO:65 is the nucleic acid sequence of the TbF (TbRa3, 38kD, Tb38-1) fusion protein.

SEQ ID NO:66 is the predicted amino acid sequence of the TbF fusion protein.

SEQ ID NO:67 is the nucleic acid sequence of the TbF2 (TbRa3, 38kD, Tb38-1, DPEP) fusion protein.

30 SEQ ID NO:68 is the predicted amino acid sequence of the TbF2 fusion protein.

SEQ ID NO:69 is the nucleic acid sequence of the TbF6 (TbRa3, 38kD, Tb38-1, TbH4) fusion protein.

SEQ ID NO:70 is the predicted amino acid sequence of the TbF6 fusion protein.

SEQ ID NO:71 is the nucleic acid sequence of the TbF8 (38kD-linker-DPEP) fusion protein.

5 SEQ ID NO:72 is the predicted amino acid sequence of the TbF8 fusion protein.

SEQ ID NO:73 is the nucleic acid sequence of the Mtb36F (Erd14-DPV-MTI) fusion protein.

10 SEQ ID NO:74 is the predicted amino acid sequence of the Mtb36F fusion protein.

SEQ ID NO:75 is the nucleic acid sequence of the Mtb88F (Erd14-DPV-MTI-MSL-MTCC#2) fusion protein.

SEQ ID NO:76 is the predicted amino acid sequence of the Mtb88F fusion protein.

15 SEQ ID NO:77 is the nucleic acid sequence of the Mtb46F (Erd14-DPV-MTI-MSL) fusion protein.

SEQ ID NO:78 is the predicted amino acid sequence of the Mtb46F fusion protein.

20 SEQ ID NO:79 is the nucleic acid sequence of the Mtb71F (DPV-MTI-MSL-MTCC#2) fusion protein.

SEQ ID NO:80 is the predicted amino acid sequence of the Mtb71F fusion protein.

SEQ ID NO:81 is the nucleic acid sequence of the Mtb31F (DPV-MTI-MSL) fusion protein.

25 SEQ ID NO:82 is the predicted amino acid sequence of the Mtb31F fusion protein.

SEQ ID NO:83 is the nucleic acid sequence of the Mtb61F (TbH9-DPV-MTI) fusion protein.

30 SEQ ID NO:84 is the predicted amino acid sequence of the Mtb61F fusion protein.

SEQ ID NO:85 is the nucleic acid sequence of the Ra12-DPPD (Mtb24F) fusion protein.

SEQ ID NO:86 is the predicted amino acid sequence of the Ra12-DPPD fusion protein.

SEQ ID NO:87 is the nucleic acid sequence of the Mtb72F (TbRa12-TbH9-TbRa35) fusion protein.

SEQ ID NO:88 is the predicted amino acid sequence of the Mtb72F fusion protein.

5 SEQ ID NO:89 is the nucleic acid sequence of the Mtb59F (TbH9-TbRa35) fusion protein.

SEQ ID NO:90 is the predicted amino acid sequence of the Mtb59F fusion protein.

SEQ ID NO:91 is the nucleic acid sequence of a vector encoding TbF14.

10 SEQ ID NO:92 is the nucleotide sequence of the region spanning nucleotides 5072 to 5095 of SEQ ID NO:91 encoding the eight amino acid His tag.

SEQ ID NO:93 is the nucleic acid sequence of a vector encoding TbF15.

SEQ ID NO:94-123 are the nucleic acid sequences of 30 overlapping peptides of HTCC#1 used for the T-cell epitope mapping.

15 SEQ ID NO:124-153 are the predicted amino acid sequences of 30 overlapping peptides of HTCC#1 used for the T-cell epitope mapping.

DETAILED DESCRIPTION OF THE INVENTION

I. INTRODUCTION

20 The present invention relates to compositions comprising antigen compositions and fusion polypeptides useful for the diagnosis and treatment of *Mycobacterium* infection, polynucleotides encoding such antigens, and methods for their use. The antigens of the present invention are polypeptides or fusion polypeptides of *Mycobacterium* antigens and immunogenic fragments thereof. More specifically, the
25 compositions of the present invention comprise at least two heterologous polypeptides of a *Mycobacterium* species of the tuberculosis complex, e.g., a species such as *M. tuberculosis*, *M. bovis*, or *M. africanum*, or a *Mycobacterium* species that is environmental or opportunistic and that causes opportunistic infections such as lung infections in immune compromised hosts (e.g., patients with AIDS), e.g., *BCG*, *M. avium*, *M. intracellulare*, *M. celatum*, *M.*
30 *genavense*, *M. haemophilum*, *M. kansasii*, *M. simiae*, *M. vaccae*, *M. fortuitum*, and *M. scrofulaceum* (see, e.g., *Harrison's Principles of Internal Medicine*, volume 1, pp. 1004-1014 and 1019-1023 (14th ed., Fauci *et al.*, eds., 1998). The inventors of the present application surprisingly discovered that compositions and fusion proteins comprising at least two

DRAFT DUE 2/10/2000

heterologous *Mycobacterium* antigens, or immunogenic fragments thereof, where highly antigenic. These compositions, fusion polypeptides, and the nucleic acids that encode them are therefore useful for eliciting protective response in patients, and for diagnostic applications.

5 The antigens of the present invention may further comprise other components designed to enhance the antigenicity of the antigens or to improve these antigens in other aspects, for example, the isolation of these antigens through addition of a stretch of histidine residues at one end of the antigen. The compositions, fusion polypeptides, and nucleic acids of the invention can comprise additional copies of antigens, or additional heterologous
10 polypeptides from *Mycobacterium* species, such as, e.g., MTb81, Mo2, TbRa3, 38 kD (with the N-terminal cysteine residue), Tb38-1, FL TbH4, HTCC#1, TbH9, MTCC#2, MTI, MSL, TbRa35, DPV, DPEP, Erd14, TbRa12, DPPD, MTb82, MTb59, ESAT-6, MTB85 complex, or α -crystalline. Such fusion polypeptides are also referred to as polyproteins. The compositions, fusion polypeptides, and nucleic acids of the invention can also comprise
15 additional polypeptides from other sources. For example, the compositions and fusion proteins of the invention can include polypeptides or nucleic acids encoding polypeptides, wherein the polypeptide enhances expression of the antigen, e.g., NS1, an influenza virus protein, or an immunogenic portion thereof (see, e.g., WO99/40188 and WO93/04175). The nucleic acids of the invention can be engineered based on codon preference in a species of
20 choice, e.g., humans.

The compositions of the invention can be naked DNA, or the compositions, e.g., polypeptides, can also comprise adjuvants such as, for example, AS2, AS2', AS2'', AS4, AS6, ENHANZYN (Detox), MPL, QS21, CWS, TDM, AGPs, CPG, Leif, saponin, and saponin mimetics, and derivatives thereof.

25 In one aspect, the compositions and fusion proteins of the invention are composed of at least two antigens selected from the group consisting of an MTb81 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and an Mo2 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex. In one embodiment, the compositions of the invention comprise the TbF14 fusion protein. The complete nucleotide sequence encoding TbF14 is set forth in SEQ ID NO:51, and the amino acid sequence of TbF14 is set forth in SEQ ID NO:52.
30

In another aspect, the compositions and fusion proteins of the invention are composed of at least four antigens selected from the group consisting of a TbRa3 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis

- complex, a 38 kD antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, a Tb38-1 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and a FL TbH4 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex.
- 5 In one embodiment, the compositions of the invention comprise the TbF15 fusion protein. The nucleic acid and amino acid sequences of TbF15 are set forth in SEQ ID NO:53 and 54, respectively.

In another aspect, the compositions and fusion proteins of the invention are composed of at least two antigens selected from the group consisting of an HTCC#1 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and a TbH9 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex. In one embodiment, the compositions of the invention comprise the HTCC#1(FL)-TbH9(FL) fusion protein. The nucleic acid and amino acid sequences of HTCC#1-TbH9 are set forth in SEQ ID NO:55 and 56, respectively. In another embodiment, the compositions of the invention comprise the fusion protein HTCC#1(184-392)/TbH9/HTCC#1(1-129). The nucleic acid and amino acid sequences of HTCC#1(184-392)/TbH9/HTCC#1(1-129) are set forth in SEQ ID NO:57 and 58, respectively. In yet another embodiment, the compositions of the invention comprise the fusion protein HTCC#1(1-149)/TbH9/HTCC#1(161-392), having the nucleic acid and amino acid sequences set forth in SEQ ID NO:59 and 60, respectively. In still another embodiment, the compositions of the invention comprise the fusion protein HTCC#1(184-392)/TbH9/HTCC#1(1-200), having the nucleic acid and amino acid sequences set forth in SEQ ID NO:61 and 62, respectively.

In a different aspect, the compositions and fusion proteins of the invention are composed of at least two antigens selected from the group consisting of an HTCC#1 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and a TbRa12 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex. In one embodiment, the compositions of the invention comprise the fusion protein TbRa12-HTCC#1. The nucleic acid and amino acid sequences of the TbRa12-HTCC#1 fusion protein are set forth in SEQ ID NO:63 and 64, respectively.

In yet another aspect, the compositions and fusion proteins of the invention are composed of at least two antigens selected from the group consisting of a TbH9 (MTB39) antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and a TbRa35 (MTB32A) antigen or an immunogenic fragment thereof

DOCUMENT ID: 20230727-141000

from a *Mycobacterium* species of the tuberculosis complex. In one embodiment, the antigens are selected from the group consisting of a TbH9 (MTB39) antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and a polypeptide comprising at least 205 amino acids of the N-terminus of a TbRa35 (MTB32A) 5 antigen from a *Mycobacterium* species of the tuberculosis complex. In another embodiment, the antigens are selected from the group consisting of a TbH9 (MTB39) antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, a polypeptide comprising at least 205 amino acids of the N-terminus of a TbRa35 (MTB32A) 10 antigen from a *Mycobacterium* species of the tuberculosis complex, and a polypeptide comprising at least about 132 amino acids from the C-terminus of a TbRa35 (MTB32A) antigen from a *Mycobacterium* species of the tuberculosis complex.

In yet another embodiment, the compositions of the invention comprise the Mtb59F fusion protein. The nucleic acid and amino acid sequences of the Mtb59F fusion protein are set forth in SEQ ID NO:89 and 90, respectively, as well as in the U.S. patent 15 application No. 09/287,849 and in the PCT/US99/07717 application. In another embodiment, the compositions of the invention comprise the Mtb72F fusion protein having the nucleic acid and amino acid sequences set forth in SEQ ID NO:87 and 88, respectively. The Mtb72F fusion protein is also disclosed in the U.S. patent application Nos. 09/223,040 and 09/223,040; and in the PCT/US99/07717 application.

20 In yet another aspect, the compositions and fusion proteins of the invention comprise at least two antigens selected from the group consisting of MTb81, Mo2, TbRa3, 38kD, Tb38-1 (MTb11), FL TbH4, HTCC#1 (Mtb40), TbH9, MTCC#2 (Mtb41), DPEP, DPPD, TbRa35, TbRa12, MTb59, MTb82, Erd14 (Mtb16), FL TbRa35 (Mtb32A), DPV (Mtb8.4), MSL (Mtb9.8), MTI (Mtb9.9A, also known as MTI-A), ESAT-6, α -crystalline, and 25 85 complex, or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

In another aspect, the fusion proteins of the invention are:

TbRa3-38 kD-Tb38-1 (TbF), the sequence of which is disclosed in SEQ ID 30 NO:65 (DNA) and SEQ ID NO:66 (protein), as well as in the U.S. patent application Nos. 08/818,112; 08/818,111; and 09/056,556; and in the WO98/16646 and WO98/16645 applications;

TbRa3-38kD-Tb38-1-DPEP (TbF2), the sequence of which is disclosed in SEQ ID NO:67 (DNA) and SEQ ID NO:68 (protein), and in the U.S. patent application Nos. 08/942,578; 08/942,341; 09/056,556; and in the WO98/16646 and WO98/16645 applications;

TbRa3-38kD-Tb38-1-TBH4 (TbF6), the sequence of which is disclosed in

5 SEQ ID NO:69 (DNA) and SEQ ID NO:70 (protein) in the U.S. patent application Nos. 08/072,967; 09/072,596; and in the PCT/US99/03268 and PCT/US99/03265 applications;

38kD-Linker-DPEP (TbF8), the sequence of which is disclosed in SEQ ID NO:71 (DNA) and SEQ ID NO:72 (protein), and in the U.S. patent application Nos. 09/072,967 and 09/072,596; as well as in the PCT/US99/03268 and PCT/US99/03265

10 applications;

Erd14-DPV-MTI (MTb36F), the sequence of which is disclosed in SEQ ID NO:73 (DNA), SEQ ID NO:74 (protein), as well as in the U.S. patent application Nos. 09/223,040 and No. 09/287,849; and in the PCT/US99/07717 application;

Erd14-DPV-MTI-MSL-MTCC#2 (MTb88f), the sequence of which is

15 disclosed in SEQ ID NO:75 (cDNA) and SEQ ID NO:76 (protein), as well as in the U.S. patent application No. 09/287,849 and in the PCT/US99/07717 application;

Erd14-DPV-MTI-MSL (MTb46F), the sequence of which is disclosed in SEQ ID NO:77 (cDNA) and SEQ ID NO:78 (protein), and in the U.S. patent application No. 09/287,849 and in the PCT/US99/07717 application;

20 DPV-MTI-MSL-MTCC#2 (MTb71F), the sequence of which is disclosed in SEQ ID NO:79 (cDNA) and SEQ ID NO:80 (protein), as well as in the U.S. patent application No. 09/287,849 and in the PCT/US99/07717 application;

DPV-MTI-MSL (MTb31F), the sequence of which is disclosed in SEQ ID NO:81 (cDNA) and SEQ ID NO:82 (protein), and in the U.S. patent application No.

25 09/287,849 and in the PCT/US99/07717 application;

TbH9-DPV-MTI (MTb61F), the sequence of which is disclosed in SEQ ID NO:83 (cDNA) and SEQ ID NO:84 (protein) (*see, also,* U.S. patent application No. 09/287,849 and PCT/US99/07717 application);

Ra12-DPPD (MTb24F), the sequence of which is disclosed in SEQ ID NO:85

30 (cDNA) and SEQ ID NO:86 (protein), as well as in the U.S. patent application No. 09/287,849 and in the PCT/US99/07717 application.

In the nomenclature of the application, TbRa35 refers to the N-terminus of MTB32A (TbRa35FL), comprising at least about the first 205 amino acids of MTB32A from *M. tuberculosis*, or the corresponding region from another *Mycobacterium* species. TbRa12

refers to the C-terminus of MTB32A (TbRa35FL), comprising at least about the last 132 amino acids from MTB32A from *M. tuberculosis*, or the corresponding region from another *Mycobacterium* species.

5 The following provides sequences of some individual antigens used in the compositions and fusion proteins of the invention:

Mtb81, the sequence of which is disclosed in SEQ ID NO:1 (DNA) and SEQ ID NO:2 (predicted amino acid).

Mo2, the sequence of which is disclosed in SEQ ID NO:3 (DNA) and SEQ ID NO:4 (predicted amino acid).

10 Tb38-1 or 38-1 (MTb11), the sequence of which is disclosed in SEQ ID NO:9 (DNA) and SEQ ID NO:10 (predicted amino acid), and is also disclosed in the U.S. patent application Nos. 09/072,96; 08/523,436; 08/523,435; 08/818,112; and 08/818,111; and in the WO97/09428 and WO97/09429 applications;

15 TbRa3, the sequence of which is disclosed in SEQ ID NO:5 (DNA) and SEQ ID NO:6 (predicted amino acid sequence) (*see, also*, WO 97/09428 and WO97/09429 applications);

20 38 kD, the sequence of which is disclosed in SEQ ID NO:7 (DNA) and SEQ ID NO:8 (predicted amino acid sequence), as well as in the U.S. patent application No. 09/072,967. 38 kD has two alternative forms, with and without the N-terminal cysteine residue;

DPEP, the sequence of which is disclosed in SEQ ID NO:39 (DNA) and SEQ ID NO:40 (predicted amino acid sequence), and in the WO97/09428 and WO97/09429 publications;

25 TbH4, the sequence of which is disclosed as SEQ ID NO:11 (DNA) and SEQ ID NO:12 (predicted amino acid sequence) (*see, also*, WO97/09428 and WO97/09429 publications);

Erd14 (MTb16), the cDNA and amino acids sequences of which are disclosed in SEQ ID NO:41 (DNA) and 42 (predicted amino acid), and in Verbon *et al.*, *J. Bacteriology* 174:1352-1359 (1992);

30 DPPD, the sequence of which is disclosed in SEQ ID NO:43 (DNA) and SEQ ID NO:44 (predicted amino acid sequence), and in the PCT/US99/03268 and PCT/US99/03265 applications. The secreted form of DPPD is shown herein in Figure 12;

MTb82 (MTb867), the sequence of which is disclosed in SEQ ID NO:47 (DNA) and SEQ ID NO:48 (predicted amino acid sequence), and in Figures 8 (DNA) and 9 (amino acid);

5 MTb59 (MTb403) , the sequence of which is disclosed in SEQ ID NO:49 (DNA) and SEQ ID NO:50 (predicted amino acid sequence), and in Figures 10 (DNA) and 11 (amino acid);

TbRa35 FL (MTB32A), the sequence of which is disclosed as SEQ ID NO:29 (cDNA) and SEQ ID NO:30 (protein), and in the U.S. patent application Nos. 08/523,436, 08/523,435; 08/658,800; 08/659,683; 08/818,112; 09/056,556; and 08/818,111; as well as in 10 the WO97/09428 and WO97/09429 applications; *see also* Skeiky *et al.*, *Infection and Immunity* 67:3998-4007 (1999);

15 TbRa12, the C-terminus of MTB32A (TbRa35FL), comprising at least about the last 132 amino acids from MTB32A from *M. tuberculosis*, the sequence of which is disclosed as SEQ ID NO:27 (DNA) and SEQ ID NO:28 (predicted amino acid sequence) (see, also, U.S. patent application No. 09/072,967; and WO97/09428 and WO97/09429 publications);

20 TbRa35, the N-terminus of MTB32A (TbRa35FL), comprising at least about the first 205 amino acids of MTB32A from *M. tuberculosis*, the nucleotide and amino acid sequence of which is disclosed in Figure 4;

TbH9 (MTB39), the sequence of which is disclosed in SEQ ID NO:25 (cDNA full length) and SEQ ID NO:26 (protein full length), as well as in the U.S. patent application Nos. 08/658,800; 08/659,683; 08/818,112; 08/818,111; and 09/056,559; and in the WO97/09428 and WO97/09429 applications.

25 HTCC#1 (MTB40), the sequence of which is disclosed in SEQ ID NO:13 (DNA) and SEQ ID NO:14 (amino acid), as well as in the U.S. patent application Nos. 09/073,010; and 09/073,009; and in the PCT/US98/10407 and PCT/US98/10514 applications;

30 MTCC#2 (MTB41), the sequence of which is disclosed in SEQ ID NO:31 (DNA) and SEQ ID NO:32 (amino acid), as well as in the U.S. patent application Nos. 09/073,010; and 09/073,009; and in the WO98/53075 and WO98/53076 publications;

MTI (Mtb9.9A), the sequence of which is disclosed in SEQ ID NO:33 (DNA) and SEQ ID NO:34 (amino acid), as well as in the U.S. patent application Nos. 09/073,010; and 09/073,009; and in the WO98/53075 and WO98/53076 publications;

MSL (Mtb9.8), the sequence of which is disclosed in SEQ ID NO:35 (DNA) and SEQ ID NO:36 (amino acid), as well as in the U.S. patent application Nos. 09/073,010; and 09/073,009; and in the WO98/53075 and WO98/53076 publications;

DPV (Mtb8.4), the sequence of which is disclosed in SEQ ID NO:37 (DNA)
5 and SEQ ID NO:38 (amino acid), and in the U.S. patent application Nos. 08/658,800; 08/659,683; 08/818,111; 08/818,112; as well as in the WO97/09428 and WO97/09429 publications;

ESAT-6 (Mtb8.4), the sequence of which is disclosed in SEQ ID NO:45 (DNA) and SEQ ID NO:46 (amino acid), and in the U.S. patent application Nos. 08/658,800;
10 08/659,683; 08/818,111; 08/818,112; as well as in the WO97/09428 and WO97/09429 publications;

The following provides sequences of some additional antigens used in the compositions and fusion proteins of the invention:

α -crystalline antigen, the sequence of which is disclosed in Verbon *et al.*, *J. Bact.* 174:1352-1359 (1992);
15 85 complex antigen, the sequence of which is disclosed in Content *et al.*, *Infect. & Immunol.* 59:3205-3212 (1991).

Each of the above sequences is also disclosed in Cole *et al.* *Nature* 393:537 (1998) and can be found at, e.g., <http://www.sanger.ac.uk> and <http://www.pasteur.fr/mycodb/>.

20 The above sequences are disclosed in U.S. patent applications Nos. 08/523,435; 08/523,436; 08/658,800; 08/659,683; 08/818,111; 08/818,112; 08/942,341; 08/942,578; 08/858,998; 08/859,381; 09/056,556; 09/072,596; 09/072,967; 09/073,009; 09/073,010; 09/223,040; 09/287,849; and in PCT patent applications PCT/US99/03265, PCT/US99/03268; PCT/US99/07717; WO97/09428; WO97/09429; WO98/16645; 25 WO98/16646; WO98/53075; and WO98/53076, each of which is herein incorporated by reference.

30 The antigens described herein include polymorphic variants and conservatively modified variations, as well as inter-strain and interspecies *Mycobacterium* homologs. In addition, the antigens described herein include subsequences or truncated sequences. The fusion proteins may also contain additional polypeptides, optionally heterologous peptides from *Mycobacterium* or other sources. These antigens may be modified, for example, by adding linker peptide sequences as described below. These linker peptides may be inserted between one or more polypeptides which make up each of the fusion proteins.

II. DEFINITIONS

“Fusion polypeptide” or “fusion protein” refers to a protein having at least two heterologous *Mycobacterium* sp. polypeptides covalently linked, either directly or via an amino acid linker. The polypeptides forming the fusion protein are typically linked C-

5 terminus to N-terminus, although they can also be linked C-terminus to C-terminus, N-terminus to N-terminus, or N-terminus to C-terminus. The polypeptides of the fusion protein can be in any order. This term also refers to conservatively modified variants, polymorphic variants, alleles, mutants, subsequences, and interspecies homologs of the antigens that make up the fusion protein. *Mycobacterium tuberculosis* antigens are described in Cole *et al.*,

10 *Nature* 393:537 (1998), which discloses the entire *Mycobacterium tuberculosis* genome. The complete sequence of *Mycobacterium tuberculosis* can also be found at

<http://www.sanger.ac.uk> and at <http://www.pasteur.fr/mycodb/> (MycDB). Antigens from other *Mycobacterium* species that correspond to *M. tuberculosis* antigens can be identified, *e.g.*, using sequence comparison algorithms, as described herein, or other methods known to those

15 of skill in the art, *e.g.*, hybridization assays and antibody binding assays.

The term “TbF14” refers to a fusion protein having at least two antigenic, heterologous polypeptides from *Mycobacterium* fused together. The two peptides are referred to as MTb81 and Mo2. This term also refers to a fusion protein having polymorphic variants, alleles, mutants, fragments, and interspecies homologs of MTb81

20 and Mo2. A nucleic acid encoding TbF14 specifically hybridizes under highly stringent hybridization conditions to SEQ ID NO:1 and 3, which individually encode the MTb81 and Mo2 antigens, respectively, and alleles, polymorphic variants, interspecies homologs, subsequences, and conservatively modified variants thereof. A TbF14 fusion polypeptide specifically binds to antibodies raised against MTb81 and Mo2, and alleles, polymorphic

25 variants, interspecies homologs, subsequences, and conservatively modified variants thereof (optionally including an amino acid linker). The antibodies are polyclonal or monoclonal.

Optional, the TbF14 fusion polypeptide specifically binds to antibodies raised against the fusion junction of MTb81 and Mo2, which antibodies do not bind to MTb81 or Mo2 individually, *i.e.*, when they are not part of a fusion protein. The individual polypeptides

30 of the fusion protein can be in any order. In some embodiments, the individual polypeptides are in order (N- to C- terminus) from large to small. Large antigens are approximately 30 to 150 kD in size, medium antigens are approximately 10 to 30 kD in

size, and small antigens are approximately less than 10 kD in size. The sequence encoding the individual polypeptide may be, *e.g.*, a fragment such as an individual CTL epitope encoding about 8 to 9 amino acids. The fragment may also include multiple epitopes. The fragment may also represent a larger part of the antigen sequence, *e.g.*, about 50% or more of MTb81 and Mo2.

TbF14 optionally comprises additional polypeptides, optionally heterologous polypeptides, fused to MTb81 and Mo2, optionally derived from *Mycobacterium* as well as other sources, such as viral, bacterial, eukaryotic, invertebrate, vertebrate, and mammalian sources. As described herein, the fusion protein can also be linked to other molecules, including additional polypeptides.

The term “TbF15” refers to a fusion protein having at least four antigenic, heterologous polypeptides from *Mycobacterium* fused together. The four peptides are referred to as TbRa3, 38 kD, Tb38-1 (with the N-terminal cysteine), and FL TbH4. This term also refers to a fusion protein having polymorphic variants, alleles, mutants, and interspecies homologs of TbRa3, 38 kD, Tb38-1, and FL TbH4. A nucleic acid encoding TbF15 specifically hybridizes under highly stringent hybridization conditions to SEQ ID NO:5, 7, 9 and 11, individually encoding TbRa3, 38 kD, Tb38-1 and FL TbH4, respectively, and alleles, fragments, polymorphic variants, interspecies homologs, subsequences, and conservatively modified variants thereof. A TbF15 fusion polypeptide specifically binds to antibodies raised against TbRa3, 38 kD, Tb38-1, and FL TbH4 and alleles, polymorphic variants, interspecies homologs, subsequences, and conservatively modified variants thereof (optionally including an amino acid linker). The antibodies are polyclonal or monoclonal. Optionally, the TbF15 fusion polypeptide specifically binds to antibodies raised against the fusion junction of TbRa3, 38 kD, Tb38-1, and FL TbH4, which antibodies do not bind to TbRa3, 38 kD, Tb38-1, and FL TbH4 individually, *i.e.*, when they are not part of a fusion protein. The polypeptides of the fusion protein can be in any order. In some embodiments, the individual polypeptides are in order (N- to C- terminus) from large to small. Large antigens are approximately 30 to 150 kD in size, medium antigens are approximately 10 to 30 kD in size, and small antigens are approximately less than 10 kD in size. The sequence encoding the individual polypeptide may be as small as, *e.g.*, a fragment such as an individual CTL epitope encoding about 8 to 9 amino acids. The fragment may also include multiple epitopes. The fragment may also represent a larger

part of the antigen sequence, *e.g.*, about 50% or more of TbRa3, 38 kD, Tb38-1, and FL TbH4.

TbF15 optionally comprises additional polypeptides, optionally heterologous polypeptides, fused to TbRa3, 38 kD, Tb38-1, and FL TbH4, optionally derived from 5 *Mycobacterium* as well as other sources such as viral, bacterial, eukaryotic, invertebrate, vertebrate, and mammalian sources. As described herein, the fusion protein can also be linked to other molecules, including additional polypeptides. The compositions of the invention can also comprise additional polypeptides that are unlinked to the fusion proteins of the invention. These additional polypeptides may be heterologous or homologous 10 polypeptides.

The “HTCC#1(Fl)-TbH9(Fl),” “HTCC#1(184-392)/TbH9/HTCC#1(1-129),” “HTCC#1(1-149)/TbH9/HTCC#1(161-392),” and “HTCC#1(184-392)/TbH9/HTCC#1(1-200)” fusion proteins refer to fusion proteins comprising at least two antigenic, heterologous polypeptides from *Mycobacterium* fused together. The two peptides 15 are referred to as HTCC#1 and TbH9. This term also refers to fusion proteins having polymorphic variants, alleles, mutants, and interspecies homologs of HTCC#1 and TbH9. A nucleic acid encoding HTCC#1-TbH9, HTCC#1(184-392)/TbH9/HTCC#1(1-129), HTCC#1(1-149)/TbH9/HTCC#1(161-392), or HTCC#1(184-392)/TbH9/HTCC#1(1-200) 20 specifically hybridizes under highly stringent hybridization conditions to SEQ ID NO:13 and 25, individually encoding HTCC#1 and TbH9, respectively, and alleles, fragments, polymorphic variants, interspecies homologs, subsequences, and conservatively modified variants thereof. A HTCC#1(Fl)-TbH9(Fl), HTCC#1(184-392)/TbH9/HTCC#1(1-129), HTCC#1(1-149)/TbH9/HTCC#1(161-392), or HTCC#1(184-392)/TbH9/HTCC#1(1-200) 25 fusion polypeptide specifically binds to antibodies raised against HTCC#1 and TbH9, and alleles, polymorphic variants, interspecies homologs, subsequences, and conservatively modified variants thereof (optionally including an amino acid linker). The antibodies are polyclonal or monoclonal. Optionally, the HTCC#1(Fl)-TbH9(Fl), HTCC#1(184-392)/TbH9/HTCC#1(1-129), HTCC#1(1-149)/TbH9/HTCC#1(161-392), or HTCC#1(184-392)/TbH9/HTCC#1(1-200) 30 fusion polypeptide specifically binds to antibodies raised against the fusion junction of the antigens, which antibodies do not bind to the antigens individually, *i.e.*, when they are not part of a fusion protein. The polypeptides of the fusion protein can be in any order. In some embodiments, the individual polypeptides are in order

(N- to C- terminus) from large to small. Large antigens are approximately 30 to 150 kD in size, medium antigens are approximately 10 to 30 kD in size, and small antigens are approximately less than 10 kD in size. The sequence encoding the individual polypeptide may be as small as, *e.g.*, a fragment such as an individual CTL epitope encoding about 8 to 5 9 amino acids. The fragment may also include multiple epitopes. The fragment may also represent a larger part of the antigen sequence, *e.g.*, about 50% or more (*e.g.*, full-length) of HTCC#1 and TbH9.

HTCC#1(FL)-TbH9(FL), HTCC#1(184-392)/TbH9/HTCC#1(1-129),
HTCC#1(1-149)/TbH9/HTCC#1(161-392), and HTCC#1(184-392)/TbH9/HTCC#1(1-200)

10 optionally comprise additional polypeptides, optionally heterologous polypeptides, fused to HTCC#1 and TbH9, optionally derived from *Mycobacterium* as well as other sources such as viral, bacterial, eukaryotic, invertebrate, vertebrate, and mammalian sources. As described herein, the fusion protein can also be linked to other molecules, including additional polypeptides. The compositions of the invention can also comprise additional 15 polypeptides that are unlinked to the fusion proteins of the invention. These additional polypeptides may be heterologous or homologous polypeptides.

The term “TbRa12-HTCC#1” refers to a fusion protein having at least two antigenic, heterologous polypeptides from *Mycobacterium* fused together. The two peptides are referred to as TbRa12 and HTCC#1. This term also refers to a fusion protein having 20 polymorphic variants, alleles, mutants, and interspecies homologs of TbRa12 and HTCC#1. A nucleic acid encoding “TbRa12-HTCC#1” specifically hybridizes under highly stringent hybridization conditions to SEQ ID NO:27 and 13, individually encoding TbRa12 and HTCC#1, respectively, and alleles, fragments, polymorphic variants, interspecies 25 homologs, subsequences, and conservatively modified variants thereof. A “TbRa12-HTCC#1” fusion polypeptide specifically binds to antibodies raised against TbRa12 and HTCC#1 and alleles, polymorphic variants, interspecies homologs, subsequences, and conservatively modified variants thereof (optionally including an amino acid linker). The 30 antibodies are polyclonal or monoclonal. Optionally, the “TbRa12-HTCC#1” fusion polypeptide specifically binds to antibodies raised against the fusion junction of TbRa12 and HTCC#1, which antibodies do not bind to TbRa12 and HTCC#1 individually, *i.e.*, when they are not part of a fusion protein. The polypeptides of the fusion protein can be in any order. In some embodiments, the individual polypeptides are in order (N- to C- terminus)

from large to small. Large antigens are approximately 30 to 150 kD in size, medium antigens are approximately 10 to 30 kD in size, and small antigens are approximately less than 10 kD in size. The sequence encoding the individual polypeptide may be as small as, e.g., a fragment such as an individual CTL epitope encoding about 8 to 9 amino acids. The 5 fragment may also include multiple epitopes. The fragment may also represent a larger part of the antigen sequence, e.g., about 50% or more of TbRa12 and HTCC#1.

“TbRa12-HTCC#1” optionally comprises additional polypeptides, optionally heterologous polypeptides, fused to TbRa12 and HTCC#1, optionally derived from *Mycobacterium* as well as other sources such as viral, bacterial, eukaryotic, invertebrate, 10 vertebrate, and mammalian sources. As described herein, the fusion protein can also be linked to other molecules, including additional polypeptides. The compositions of the invention can also comprise additional polypeptides that are unlinked to the fusion proteins of the invention. These additional polypeptides may be heterologous or homologous polypeptides.

15 The term “Mtb72F” and “Mtb59F” refer to fusion proteins of the invention which hybridize under stringent conditions to at least two nucleotide sequences set forth in SEQ ID NO:25 and 29, individually encoding the TbH9 (MTB39) and Ra35 (MTB32A) antigens. The polynucleotide sequences encoding the individual antigens of the fusion polypeptides therefore include conservatively modified variants, polymorphic variants, 20 alleles, mutants, subsequences, and interspecies homologs of TbH9 (MTB39) and Ra35 (MTB32A). The polynucleotide sequence encoding the individual polypeptides of the fusion proteins can be in any order. In some embodiments, the individual polypeptides are in order (N- to C- terminus) from large to small. Large antigens are approximately 30 to 150 kD in size, medium antigens are approximately 10 to 30 kD in size, and small antigens are 25 approximately less than 10 kD in size. The sequence encoding the individual polypeptide may be as small as, e.g., a fragment such as an individual CTL epitope encoding about 8 to 9 amino acids. The fragment may also include multiple epitopes. The fragment may also represent a larger part of the antigen sequence, e.g., about 50% or more of TbH9 (MTB39) and Ra35 (MTB32A), e.g., the N- and C-terminal portions of Ra35 (MTB32A).

30 An “Mtb72F” or “Mtb59F” fusion polypeptide of the invention specifically binds to antibodies raised against at least two antigen polypeptides, wherein each antigen polypeptide is selected from the group consisting of TbH9 (MTB39) and Ra35 (MTB32A). The antibodies can be polyclonal or monoclonal. Optionally, the fusion polypeptide

specifically binds to antibodies raised against the fusion junction of the antigens, which antibodies do not bind to the antigens individually, *i.e.*, when they are not part of a fusion protein. The fusion polypeptides optionally comprise additional polypeptides, *e.g.*, three, four, five, six, or more polypeptides, up to about 25 polypeptides, optionally heterologous polypeptides or repeated homologous polypeptides, fused to the at least two heterologous antigens. The additional polypeptides of the fusion protein are optionally derived from *Mycobacterium* as well as other sources, such as other bacterial, viral, or invertebrate, vertebrate, or mammalian sources. The individual polypeptides of the fusion protein can be in any order. As described herein, the fusion protein can also be linked to other molecules, including additional polypeptides. The compositions of the invention can also comprise additional polypeptides that are unlinked to the fusion proteins of the invention. These additional polypeptides may be heterologous or homologous polypeptides.

A polynucleotide sequence comprising a fusion protein of the invention hybridizes under stringent conditions to at least two nucleotide sequences, each encoding an antigen polypeptide selected from the group consisting of Mtb81, Mo2, TbRa3, 38 kD, Tb38-1, TbH4, HTCC#1, TbH9, MTCC#2, MTI, MSL, TbRa35, DPV, DPEP, Erd14, TbRa12, DPPD, ESAT-6, MTb82, MTb59, Mtb85 complex, and α -crystalline. The polynucleotide sequences encoding the individual antigens of the fusion polypeptide therefore include conservatively modified variants, polymorphic variants, alleles, mutants, subsequences, and interspecies homologs of Mtb81, Mo2, TbRa3, 38 kD, Tb38-1, TbH4, HTCC#1, TbH9, MTCC#2, MTI, MSL, TbRa35, DPV, DPEP, Erd14, TbRa12, DPPD, ESAT-6, MTb82, MTb59, Mtb85 complex, and α -crystalline. The polynucleotide sequence encoding the individual polypeptides of the fusion protein can be in any order. In some embodiments, the individual polypeptides are in order (N- to C- terminus) from large to small. Large antigens are approximately 30 to 150 kD in size, medium antigens are approximately 10 to 30 kD in size, and small antigens are approximately less than 10 kD in size. The sequence encoding the individual polypeptide may be as small as, *e.g.*, a fragment such as an individual CTL epitope encoding about 8 to 9 amino acids. The fragment may also include multiple epitopes. The fragment may also represent a larger part of the antigen sequence, *e.g.*, about 50% or more of Mtb81, Mo2, TbRa3, 38 kD, Tb38-1, TbH4, HTCC#1, TbH9, MTCC#2, MTI, MSL, TbRa35, DPV, DPEP, Erd14, TbRa12, DPPD, ESAT-6, MTb82, MTb59, Mtb85 complex, and α -crystalline.

A fusion polypeptide of the invention specifically binds to antibodies raised against at least two antigen polypeptides, wherein each antigen polypeptide is selected from the group consisting of Mtb81, Mo2, TbRa3, 38 kD, Tb38-1, TbH4, HTCC#1, TbH9, MTCC#2, MTI, MSL, TbRa35, DPV, DPEP, Erd14, TbRa12, DPPD, ESAT-6, MTb82, 5 MTb59, Mtb85 complex, and α -crystalline. The antibodies can be polyclonal or monoclonal. Optionally, the fusion polypeptide specifically binds to antibodies raised against the fusion junction of the antigens, which antibodies do not bind to the antigens individually, *i.e.*, when they are not part of a fusion protein. The fusion polypeptides optionally comprise additional polypeptides, *e.g.*, three, four, five, six, or more polypeptides, up to about 25 10 polypeptides, optionally heterologous polypeptides or repeated homologous polypeptides, fused to the at least two heterologous antigens. The additional polypeptides of the fusion protein are optionally derived from *Mycobacterium* as well as other sources, such as other bacterial, viral, or invertebrate, vertebrate, or mammalian sources. The individual 15 polypeptides of the fusion protein can be in any order. As described herein, the fusion protein can also be linked to other molecules, including additional polypeptides. The compositions of the invention can also comprise additional polypeptides that are unlinked to the fusion proteins of the invention. These additional polypeptides may be heterologous or homologous polypeptides.

The term “fused” refers to the covalent linkage between two polypeptides in a 20 fusion protein. The polypeptides are typically joined via a peptide bond, either directly to each other or via an amino acid linker. Optionally, the peptides can be joined via non-peptide covalent linkages known to those of skill in the art.

“FL” refers to full-length, *i.e.*, a polypeptide that is the same length as the wild-type polypeptide.

25 The term “immunogenic fragment thereof” refers to a polypeptide comprising an epitope that is recognized by cytotoxic T lymphocytes, helper T lymphocytes or B cells.

The term “*Mycobacterium* species of the tuberculosis complex” includes those species traditionally considered as causing the disease tuberculosis, as well as *Mycobacterium* environmental and opportunistic species that cause tuberculosis and lung disease in immune 30 compromised patients, such as patients with AIDS, *e.g.*, *M. tuberculosis*, *M. bovis*, or *M. africanum*, *BCG*, *M. avium*, *M. intracellulare*, *M. celatum*, *M. genavense*, *M. haemophilum*, *M. kansasii*, *M. simiae*, *M. vaccae*, *M. fortuitum*, and *M. scrofulaceum* (*see, e.g., Harrison’s*

Principles of Internal Medicine, volume 1, pp. 1004-1014 and 1019-1023 (14th ed., Fauci *et al.*, eds., 1998).

An adjuvant refers to the components in a vaccine or therapeutic composition that increase the specific immune response to the antigen (*see, e.g.*, Edelman, *AIDS Res. Hum Retroviruses* 8:1409-1411 (1992)). Adjuvants induce immune responses of the Th1-type and Th-2 type response. Th1-type cytokines (*e.g.*, IFN- γ , IL-2, and IL-12) tend to favor the induction of cell-mediated immune response to an administered antigen, while Th-2 type cytokines (*e.g.*, IL-4, IL-5, IL-6, IL-10 and TNF- β) tend to favor the induction of humoral immune responses.

“Nucleic acid” refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. The term encompasses nucleic acids containing known nucleotide analogs or modified backbone residues or linkages, which are synthetic, naturally occurring, and non-naturally occurring, which have similar binding properties as the reference nucleic acid, and which are metabolized in a manner similar to the reference nucleotides. Examples of such analogs include, without limitation, phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2-O-methyl ribonucleotides, peptide-nucleic acids (PNAs).

Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (*e.g.*, degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al.*, *Nucleic Acid Res.* 19:5081 (1991); Ohtsuka *et al.*, *J. Biol. Chem.* 260:2605-2608 (1985); Rossolini *et al.*, *Mol. Cell. Probes* 8:91-98 (1994)). The term nucleic acid is used interchangeably with gene, cDNA, mRNA, oligonucleotide, and polynucleotide.

The terms “polypeptide,” “peptide” and “protein” are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymer.

The term “amino acid” refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to

the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, *e.g.*, hydroxyproline, γ -carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, *i.e.*, an α carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, *e.g.*, homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (*e.g.*, norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions in a manner similar to a naturally occurring amino acid.

Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

“Conservatively modified variants” applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are “silent variations,” which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a “conservatively modified variant” where the alteration

results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the invention.

5 The following eight groups each contain amino acids that are conservative substitutions for one another:

- 1) Alanine (A), Glycine (G);
- 2) Aspartic acid (D), Glutamic acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 10 4) Arginine (R), Lysine (K);
- 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V);
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W);
- 7) Serine (S), Threonine (T); and
- 8) Cysteine (C), Methionine (M)

15 (see, e.g., Creighton, *Proteins* (1984)).

The term “heterologous” when used with reference to portions of a nucleic acid indicates that the nucleic acid comprises two or more subsequences that are not found in the same relationship to each other in nature. For instance, the nucleic acid is typically recombinantly produced, having two or more sequences from unrelated genes arranged to make a new functional nucleic acid, e.g., a promoter from one source and a coding region from another source. Similarly, a heterologous protein indicates that the protein comprises two or more subsequences that are not found in the same relationship to each other in nature (e.g., a fusion protein).

20 The phrase “selectively (or specifically) hybridizes to” refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent hybridization conditions when that sequence is present in a complex mixture (e.g., total cellular or library DNA or RNA).

25 The phrase “stringent hybridization conditions” refers to conditions under which a probe will hybridize to its target subsequence, typically in a complex mixture of nucleic acid, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Probes*, “Overview of principles of hybridization and the strategy of nucleic acid assays”

(1993). Generally, stringent conditions are selected to be about 5-10°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength pH. The T_m is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m , 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (e.g., 10 to 50 nucleotides) and at least about 60°C for long probes (e.g., greater than 50 nucleotides).

- 5 Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal is at least two times background, optionally 10 times background hybridization. Exemplary stringent hybridization conditions can be as following: 50% formamide, 5x SSC, and 1% SDS, incubating at 42°C, or, 5x SSC, 1% SDS, incubating at 65°C, with wash in 0.2x SSC, and 15 0.1% SDS at 65°C.

Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, for example, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. In such cases, the nucleic acids typically hybridize 20 under moderately stringent hybridization conditions. Exemplary “moderately stringent hybridization conditions” include a hybridization in a buffer of 40% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 1X SSC at 45°C. A positive hybridization is at least twice background. Those of ordinary skill will readily recognize that alternative hybridization and wash conditions can be utilized to provide conditions of similar stringency.

- 25 “Antibody” refers to a polypeptide comprising a framework region from an immunoglobulin gene or fragments thereof that specifically binds and recognizes an antigen. The recognized immunoglobulin genes include the kappa, lambda, alpha, gamma, delta, epsilon, and mu constant region genes, as well as the myriad immunoglobulin variable region genes. Light chains are classified as either kappa or lambda. Heavy chains are classified as 30 gamma, mu, alpha, delta, or epsilon, which in turn define the immunoglobulin classes, IgG, IgM, IgA, IgD and IgE, respectively.

An exemplary immunoglobulin (antibody) structural unit comprises a tetramer. Each tetramer is composed of two identical pairs of polypeptide chains, each pair having one “light” (about 25 kDa) and one “heavy” chain (about 50-70 kDa). The N-

terminus of each chain defines a variable region of about 100 to 110 or more amino acids primarily responsible for antigen recognition. The terms variable light chain (V_L) and variable heavy chain (V_H) refer to these light and heavy chains respectively.

Antibodies exist, e.g., as intact immunoglobulins or as a number of well-characterized fragments produced by digestion with various peptidases. Thus, for example, pepsin digests an antibody below the disulfide linkages in the hinge region to produce $F(ab)'_2$, a dimer of Fab which itself is a light chain joined to V_H-C_H1 by a disulfide bond. The $F(ab)'_2$ may be reduced under mild conditions to break the disulfide linkage in the hinge region, thereby converting the $F(ab)'_2$ dimer into an Fab' monomer. The Fab' monomer is essentially Fab with part of the hinge region (see *Fundamental Immunology* (Paul ed., 3d ed. 1993). While various antibody fragments are defined in terms of the digestion of an intact antibody, one of skill will appreciate that such fragments may be synthesized *de novo* either chemically or by using recombinant DNA methodology. Thus, the term antibody, as used herein, also includes antibody fragments either produced by the modification of whole antibodies, or those synthesized *de novo* using recombinant DNA methodologies (e.g., single chain Fv) or those identified using phage display libraries (see, e.g., McCafferty et al., *Nature* 348:552-554 (1990)).

For preparation of monoclonal or polyclonal antibodies, any technique known in the art can be used (see, e.g., Kohler & Milstein, *Nature* 256:495-497 (1975); Kozbor et al., *Immunology Today* 4: 72 (1983); Cole et al., pp. 77-96 in *Monoclonal Antibodies and Cancer Therapy* (1985)). Techniques for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms such as other mammals, may be used to express humanized antibodies. Alternatively, phage display technology can be used to identify antibodies and heteromeric Fab fragments that specifically bind to selected antigens (see, e.g., McCafferty et al., *Nature* 348:552-554 (1990); Marks et al., *Biotechnology* 10:779-783 (1992)).

The phrase "specifically (or selectively) binds" to an antibody or "specifically (or selectively) immunoreactive with," when referring to a protein or peptide, refers to a binding reaction that is determinative of the presence of the protein in a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bind to a particular protein at least two times the background and do not substantially bind in a significant amount to other proteins present in the sample. Specific binding to an antibody under such conditions may require an antibody that is selected for its

specificity for a particular protein. For example, polyclonal antibodies raised to fusion proteins can be selected to obtain only those polyclonal antibodies that are specifically immunoreactive with fusion protein and not with individual components of the fusion proteins. This selection may be achieved by subtracting out antibodies that cross-react with the individual antigens. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays are routinely used to select antibodies specifically immunoreactive with a protein (see, e.g., Harlow & Lane, *Antibodies, A Laboratory Manual* (1988), for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity). Typically a specific or selective reaction will be at least twice background signal or noise and more typically more than 10 to 100 times background.

Polynucleotides may comprise a native sequence (*i.e.*, an endogenous sequence that encodes an individual antigen or a portion thereof) or may comprise a variant of such a sequence. Polynucleotide variants may contain one or more substitutions, additions, deletions and/or insertions such that the biological activity of the encoded fusion polypeptide is not diminished, relative to a fusion polypeptide comprising native antigens. Variants preferably exhibit at least about 70% identity, more preferably at least about 80% identity and most preferably at least about 90% identity to a polynucleotide sequence that encodes a native polypeptide or a portion thereof.

The terms “identical” or percent “identity,” in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same (*i.e.*, 70% identity, optionally 75%, 80%, 85%, 90%, or 95% identity over a specified region), when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. Such sequences are then said to be “substantially identical.” This definition also refers to the compliment of a test sequence. Optionally, the identity exists over a region that is at least about 25 to about 50 amino acids or nucleotides in length, or optionally over a region that is 75-100 amino acids or nucleotides in length.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program

parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

- A “comparison window”, as used herein, includes reference to a segment of
- 5 any one of the number of contiguous positions selected from the group consisting of from 25 to 500, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison
- 10 can be conducted, *e.g.*, by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics
- 15 Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection (*see, e.g.*, *Current Protocols in Molecular Biology* (Ausubel *et al.*, eds. 1995 supplement)).

One example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments

20 to show relationship and percent sequence identity. It also plots a tree or dendrogram showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng & Doolittle, *J. Mol. Evol.* 35:351-360 (1987). The method used is similar to the method described by Higgins & Sharp, *CABIOS* 5:151-153 (1989). The program can align up to 300 sequences, each of a maximum length of 5,000

25 nucleotides or amino acids. The multiple alignment procedure begins with the pairwise alignment of the two most similar sequences, producing a cluster of two aligned sequences. This cluster is then aligned to the next most related sequence or cluster of aligned sequences. Two clusters of sequences are aligned by a simple extension of the pairwise alignment of two individual sequences. The final alignment is achieved by a series of progressive, pairwise

30 alignments. The program is run by designating specific sequences and their amino acid or nucleotide coordinates for regions of sequence comparison and by designating the program parameters. Using PILEUP, a reference sequence is compared to other test sequences to determine the percent sequence identity relationship using the following parameters: default gap weight (3.00), default gap length weight (0.10), and weighted end gaps. PILEUP can be

obtained from the GCG sequence analysis software package, e.g., version 7.0 (Devereaux *et al.*, *Nuc. Acids Res.* 12:387-395 (1984)).

Another example of algorithm that is suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al.*, *Nuc. Acids Res.* 25:3389-3402 (1977) and Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990), respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) or 10, M=5, N=-4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3, and expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a

comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

III. POLYNUCLEOTIDE COMPOSITIONS

As used herein, the terms "DNA segment" and "polynucleotide" refer to a
5 DNA molecule that has been isolated free of total genomic DNA of a particular species. Therefore, a DNA segment encoding a polypeptide refers to a DNA segment that contains one or more coding sequences yet is substantially isolated away from, or purified free from, total genomic DNA of the species from which the DNA segment is obtained. Included within the terms "DNA segment" and "polynucleotide" are DNA segments and smaller fragments of
10 such segments, and also recombinant vectors, including, for example, plasmids, cosmids, phagemids, phage, viruses, and the like.

As will be understood by those skilled in the art, the DNA segments of this invention can include genomic sequences, extra-genomic and plasmid-encoded sequences and smaller engineered gene segments that express, or may be adapted to express, proteins,
15 polypeptides, peptides and the like. Such segments may be naturally isolated, or modified synthetically by the hand of man.

"Isolated," as used herein, means that a polynucleotide is substantially away from other coding sequences, and that the DNA segment does not contain large portions of unrelated coding DNA, such as large chromosomal fragments or other functional genes or
20 polypeptide coding regions. Of course, this refers to the DNA segment as originally isolated, and does not exclude genes or coding regions later added to the segment by the hand of man.

As will be recognized by the skilled artisan, polynucleotides may be single-stranded (coding or antisense) or double-stranded, and may be DNA (genomic, cDNA or synthetic) or RNA molecules. RNA molecules include HnRNA molecules, which contain
25 introns and correspond to a DNA molecule in a one-to-one manner, and mRNA molecules, which do not contain introns. Additional coding or non-coding sequences may, but need not, be present within a polynucleotide of the present invention, and a polynucleotide may, but need not, be linked to other molecules and/or support materials.

Polynucleotides may comprise a native sequence (*i.e.*, an endogenous
30 sequence that encodes a *Mycobacterium* antigen or a portion thereof) or may comprise a variant, or a biological or antigenic functional equivalent of such a sequence. Polynucleotide variants may contain one or more substitutions, additions, deletions and/or insertions, as further described below, preferably such that the immunogenicity of the encoded polypeptide

is not diminished, relative to a native tumor protein. The effect on the immunogenicity of the encoded polypeptide may generally be assessed as described herein. The term "variants" also encompasses homologous genes of xenogenic origin.

In additional embodiments, the present invention provides isolated
5 polynucleotides and polypeptides comprising various lengths of contiguous stretches of sequence identical to or complementary to one or more of the sequences disclosed herein. For example, polynucleotides are provided by this invention that comprise at least about 15, 20, 30, 40, 50, 75, 100, 150, 200, 300, 400, 500 or 1000 or more contiguous nucleotides of one or more of the sequences disclosed herein as well as all intermediate lengths there
10 between. It will be readily understood that "intermediate lengths", in this context, means any length between the quoted values, such as 16, 17, 18, 19, etc.; 21, 22, 23, etc.; 30, 31, 32, etc.; 50, 51, 52, 53, etc.; 100, 101, 102, 103, etc.; 150, 151, 152, 153, etc.; including all integers through 200-500; 500-1,000, and the like.

The polynucleotides of the present invention, or fragments thereof, regardless
15 of the length of the coding sequence itself, may be combined with other DNA sequences, such as promoters, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably. It is therefore contemplated that a nucleic acid fragment of almost any length may be employed, with the total length preferably being limited by the ease of preparation
20 and use in the intended recombinant DNA protocol. For example, illustrative DNA segments with total lengths of about 10,000, about 5000, about 3000, about 2,000, about 1,000, about 500, about 200, about 100, about 50 base pairs in length, and the like, (including all intermediate lengths) are contemplated to be useful in many implementations of this invention.

25 Moreover, it will be appreciated by those of ordinary skill in the art that, as a result of the degeneracy of the genetic code, there are many nucleotide sequences that encode a polypeptide as described herein. Some of these polynucleotides bear minimal homology to the nucleotide sequence of any native gene. Nonetheless, polynucleotides that vary due to differences in codon usage are specifically contemplated by the present invention, for
30 example polynucleotides that are optimized for human and/or primate codon selection. Further, alleles of the genes comprising the polynucleotide sequences provided herein are within the scope of the present invention. Alleles are endogenous genes that are altered as a result of one or more mutations, such as deletions, additions and/or substitutions of nucleotides. The resulting mRNA and protein may, but need not, have an altered structure or

function. Alleles may be identified using standard techniques (such as hybridization, amplification and/or database sequence comparison).

IV. POLYNUCLEOTIDE IDENTIFICATION AND CHARACTERIZATION

Polynucleotides may be identified, prepared and/or manipulated using any of a variety of well established techniques. For example, a polynucleotide may be identified, as described in more detail below, by screening a microarray of cDNAs for tumor-associated expression (*i.e.*, expression that is at least two fold greater in a tumor than in normal tissue, as determined using a representative assay provided herein). Such screens may be performed, for example, using a Synteni microarray (Palo Alto, CA) according to the manufacturer's instructions (and essentially as described by Schena *et al.*, *Proc. Natl. Acad. Sci. USA* 93:10614-10619 (1996) and Heller *et al.*, *Proc. Natl. Acad. Sci. USA* 94:2150-2155 (1997)). Alternatively, polynucleotides may be amplified from cDNA prepared from cells expressing the proteins described herein, such as *M. tuberculosis* cells. Such polynucleotides may be amplified via polymerase chain reaction (PCR). For this approach, sequence-specific primers may be designed based on the sequences provided herein, and may be purchased or synthesized.

An amplified portion of a polynucleotide of the present invention may be used to isolate a full length gene from a suitable library (*e.g.*, a *M. tuberculosis* cDNA library) using well known techniques. Within such techniques, a library (cDNA or genomic) is screened using one or more polynucleotide probes or primers suitable for amplification. Preferably, a library is size-selected to include larger molecules. Random primed libraries may also be preferred for identifying 5' and upstream regions of genes. Genomic libraries are preferred for obtaining introns and extending 5' sequences.

For hybridization techniques, a partial sequence may be labeled (*e.g.*, by nick-translation or end-labeling with ^{32}P) using well known techniques. A bacterial or bacteriophage library is then generally screened by hybridizing filters containing denatured bacterial colonies (or lawns containing phage plaques) with the labeled probe (*see* Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual* (1989)). Hybridizing colonies or plaques are selected and expanded, and the DNA is isolated for further analysis. cDNA clones may be analyzed to determine the amount of additional sequence by, for example, PCR using a primer from the partial sequence and a primer from the vector. Restriction maps and partial sequences may be generated to identify one or more overlapping clones. The complete sequence may then be determined using standard techniques, which may involve generating a

series of deletion clones. The resulting overlapping sequences can then assembled into a single contiguous sequence. A full length cDNA molecule can be generated by ligating suitable fragments, using well known techniques.

Alternatively, there are numerous amplification techniques for obtaining a full length coding sequence from a partial cDNA sequence. Within such techniques, 5 amplification is generally performed via PCR. Any of a variety of commercially available kits may be used to perform the amplification step. Primers may be designed using, for example, software well known in the art. Primers are preferably 22-30 nucleotides in length, have a GC content of at least 50% and anneal to the target sequence at temperatures of about 10 68°C to 72°C. The amplified region may be sequenced as described above, and overlapping sequences assembled into a contiguous sequence.

One such amplification technique is inverse PCR (*see* Triglia *et al.*, *Nucl. Acids Res.* 16:8186 (1988)), which uses restriction enzymes to generate a fragment in the known region of the gene. The fragment is then circularized by intramolecular ligation and 15 used as a template for PCR with divergent primers derived from the known region. Within an alternative approach, sequences adjacent to a partial sequence may be retrieved by amplification with a primer to a linker sequence and a primer specific to a known region. The amplified sequences are typically subjected to a second round of amplification with the same linker primer and a second primer specific to the known region. A variation on this 20 procedure, which employs two primers that initiate extension in opposite directions from the known sequence, is described in WO 96/38591. Another such technique is known as “rapid amplification of cDNA ends” or RACE. This technique involves the use of an internal primer and an external primer, which hybridizes to a polyA region or vector sequence, to identify sequences that are 5’ and 3’ of a known sequence. Additional techniques include 25 capture PCR (Lagerstrom *et al.*, *PCR Methods Appl.* 1:111-19 (1991)) and walking PCR (Parker *et al.*, *Nucl. Acids. Res.* 19:3055-60 (1991)). Other methods employing amplification may also be employed to obtain a full length cDNA sequence.

In certain instances, it is possible to obtain a full length cDNA sequence by 30 analysis of sequences provided in an expressed sequence tag (EST) database, such as that available from GenBank. Searches for overlapping ESTs may generally be performed using well known programs (e.g., NCBI BLAST searches), and such ESTs may be used to generate a contiguous full length sequence. Full length DNA sequences may also be obtained by analysis of genomic fragments.

V. POLYNUCLEOTIDE EXPRESSION IN HOST CELLS

In other embodiments of the invention, polynucleotide sequences or fragments thereof which encode polypeptides of the invention, or fusion proteins or functional equivalents thereof, may be used in recombinant DNA molecules to direct expression of a polypeptide in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences that encode substantially the same or a functionally equivalent amino acid sequence may be produced and these sequences may be used to clone and express a given polypeptide.

As will be understood by those of skill in the art, it may be advantageous in some instances to produce polypeptide-encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce a recombinant RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence.

Moreover, the polynucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter polypeptide encoding sequences for a variety of reasons, including but not limited to, alterations which modify the cloning, processing, and/or expression of the gene product. For example, DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. In addition, site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, or introduce mutations, and so forth.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences may be ligated to a heterologous sequence to encode a fusion protein. For example, to screen peptide libraries for inhibitors of polypeptide activity, it may be useful to encode a chimeric protein that can be recognized by a commercially available antibody. A fusion protein may also be engineered to contain a cleavage site located between the polypeptide-encoding sequence and the heterologous protein sequence, so that the polypeptide may be cleaved and purified away from the heterologous moiety.

Sequences encoding a desired polypeptide may be synthesized, in whole or in part, using chemical methods well known in the art (see Caruthers, M. H. *et al.*, *Nucl. Acids Res. Symp. Ser.* pp. 215-223 (1980), Horn *et al.*, *Nucl. Acids Res. Symp. Ser.* pp. 225-232 (1980)). Alternatively, the protein itself may be produced using chemical methods to

DRAFT DRAFT DRAFT DRAFT

synthesize the amino acid sequence of a polypeptide, or a portion thereof. For example, peptide synthesis can be performed using various solid-phase techniques (Roberge *et al.*, *Science* 269:202-204 (1995)) and automated synthesis may be achieved, for example, using the ABI 431A Peptide Synthesizer (Perkin Elmer, Palo Alto, CA).

5 A newly synthesized peptide may be substantially purified by preparative high performance liquid chromatography (e.g., Creighton, *Proteins, Structures and Molecular Principles* (1983)) or other comparable techniques available in the art. The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (e.g., the Edman degradation procedure). Additionally, the amino acid sequence of a polypeptide, or
10 any part thereof, may be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

In order to express a desired polypeptide, the nucleotide sequences encoding the polypeptide, or functional equivalents, may be inserted into appropriate expression vector,
15 *i.e.*, a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence. Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding a polypeptide of interest and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination.
20 Such techniques are described in Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (1989), and Ausubel *et al.*, *Current Protocols in Molecular Biology* (1989).

A variety of expression vector/host systems may be utilized to contain and express polynucleotide sequences. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA
25 expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus); plant cell systems transformed with virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems.

The “control elements” or “regulatory sequences” present in an expression
30 vector are those non-translated regions of the vector--enhancers, promoters, 5’ and 3’ untranslated regions--which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in

bacterial systems, inducible promoters such as the hybrid lacZ promoter of the PBLUESCRIPT phagemid (Stratagene, La Jolla, Calif.) or PSSPORT1 plasmid (Gibco BRL, Gaithersburg, MD) and the like may be used. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are generally preferred. If it is necessary to
5 generate a cell line that contains multiple copies of the sequence encoding a polypeptide, vectors based on SV40 or EBV may be advantageously used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for the expressed polypeptide. For example, when large
10 quantities are needed, for example for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be used. Such vectors include, but are not limited to, the multifunctional *E. coli* cloning and expression vectors such as BLUESCRIPT (Stratagene), in which the sequence encoding the polypeptide of interest may be ligated into the vector in frame with sequences for the amino-terminal Met and the
15 subsequent 7 residues of β -galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke & Schuster, *J. Biol. Chem.* 264:5503-5509 (1989)); and the like. pGEX Vectors (Promega, Madison, Wis.) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads
20 followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast, *Saccharomyces cerevisiae*, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase, and PGH may be
25 used. For reviews, see Ausubel *et al.* (*supra*) and Grant *et al.*, *Methods Enzymol.* 153:516-544 (1987).

In cases where plant expression vectors are used, the expression of sequences encoding polypeptides may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination
30 with the omega leader sequence from TMV (Takamatsu, *EMBO J.* 6:307-311 (1987)). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi *et al.*, *EMBO J.* 3:1671-1680 (1984); Broglie *et al.*, *Science* 224:838-843 (1984); and Winter *et al.*, *Results Probl. Cell Differ.* 17:85-105 (1991)). These

constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews (see, e.g., Hobbs in *McGraw Hill Yearbook of Science and Technology* pp. 191-196 (1992)).

5 An insect system may also be used to express a polypeptide of interest. For example, in one such system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in *Spodoptera frugiperda* cells or in *Trichoplusia larvae*. The sequences encoding the polypeptide may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter.

10 Successful insertion of the polypeptide-encoding sequence will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, *S. frugiperda* cells or *Trichoplusia larvae* in which the polypeptide of interest may be expressed (Engelhard *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 91 :3224-3227 (1994)).

15 In mammalian host cells, a number of viral-based expression systems are generally available. For example, in cases where an adenovirus is used as an expression vector, sequences encoding a polypeptide of interest may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain a viable virus which is capable of expressing the polypeptide in infected host cells (Logan & Shenk, *Proc. Natl. Acad. Sci. U.S.A.* 81:3655-3659 (1984)). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

20 Specific initiation signals may also be used to achieve more efficient translation of sequences encoding a polypeptide of interest. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding the polypeptide, its initiation codon, and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a portion thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of

enhancers which are appropriate for the particular cell system which is used, such as those described in the literature (Scharf. *et al.*, *Results Probl. Cell Differ.* 20:125-162 (1994)).

In addition, a host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to facilitate correct insertion, folding and/or function. Different host cells such as CHO, HeLa, MDCK, HEK293, and WI38, which have specific cellular machinery and characteristic mechanisms for such post-translational activities, may be chosen to ensure the correct modification and processing of the foreign protein.

For long-term, high-yield production of recombinant proteins, stable expression is generally preferred. For example, cell lines which stably express a polynucleotide of interest may be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler *et al.*, *Cell* 11:223-32 (1977)) and adenine phosphoribosyltransferase (Lowy *et al.*, *Cell* 22:817-23 (1990)) genes which can be employed in tk.sup.- or aprt.sup.- cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to methotrexate (Wigler *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 77:3567-70 (1980)); npt, which confers resistance to the aminoglycosides, neomycin and G-418 (Colbere-Garapin *et al.*, *J. Mol. Biol.* 150:1-14 (1981)); and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, *supra*). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histidol in place of histidine (Hartman & Mulligan, *Proc. Natl. Acad. Sci. U.S.A.* 85:8047-51 (1988)). Recently, the use of visible markers has gained popularity with such markers as

anthocyanins, β -glucuronidase and its substrate GUS, and luciferase and its substrate luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes *et al.*, *Methods Mol. Biol.* 55:121-131 (1995)).

5 Although the presence/absence of marker gene expression suggests that the gene of interest is also present, its presence and expression may need to be confirmed. For example, if the sequence encoding a polypeptide is inserted within a marker gene sequence, recombinant cells containing sequences can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a polypeptide-encoding 10 sequence under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

15 Alternatively, host cells which contain and express a desired polynucleotide sequence may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein.

20 A variety of protocols for detecting and measuring the expression of polynucleotide-encoded products, using either polyclonal or monoclonal antibodies specific for the product are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on a given polypeptide may be preferred for some applications, but a competitive binding assay may also be employed. These and other assays are described, among other places, in Hampton *et al.*, *Serological Methods, a Laboratory Manual* (1990) 25 and Maddox *et al.*, *J. Exp. Med.* 158:1211-1216 (1983).

30 A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide. Alternatively, the sequences, or any portions thereof may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides.

These procedures may be conducted using a variety of commercially available kits. Suitable reporter molecules or labels, which may be used include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

5 Host cells transformed with a polynucleotide sequence of interest may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a recombinant cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides of the invention may
10 be designed to contain signal sequences which direct secretion of the encoded polypeptide through a prokaryotic or eukaryotic cell membrane. Other recombinant constructions may be used to join sequences encoding a polypeptide of interest to nucleotide sequence encoding a polypeptide domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-
15 tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, Washington). The inclusion of cleavable linker sequences such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, Calif.) between the purification domain and the encoded polypeptide
20 may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing a polypeptide of interest and a nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on IMIAC (immobilized metal ion affinity chromatography) as described in Porath *et al.*, *Prot. Exp. Purif.* 3:263-281 (1992) while the enterokinase cleavage
25 site provides a means for purifying the desired polypeptide from the fusion protein. A discussion of vectors which contain fusion proteins is provided in Kroll *et al.*, *DNA Cell Biol.* 12:441-453 (1993)).

In addition to recombinant production methods, polypeptides of the invention, and fragments thereof, may be produced by direct peptide synthesis using solid-phase
30 techniques (Merrifield, *J. Am. Chem. Soc.* 85:2149-2154 (1963)). Protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer). Alternatively, various fragments may be chemically synthesized separately and combined using chemical methods to produce the full length molecule.

VI. IN VIVO POLYNUCLEOTIDE DELIVERY TECHNIQUES

In additional embodiments, genetic constructs comprising one or more of the polynucleotides of the invention are introduced into cells *in vivo*. This may be achieved using any of a variety of well known approaches, several of which are outlined below for the purpose of illustration.

1. Adenovirus

One of the preferred methods for *in vivo* delivery of one or more nucleic acid sequences involves the use of an adenovirus expression vector. "Adenovirus expression vector" is meant to include those constructs containing adenovirus sequences sufficient to (a) support packaging of the construct and (b) to express a polynucleotide that has been cloned therein in a sense or antisense orientation. Of course, in the context of an antisense construct, expression does not require that the gene product be synthesized.

The expression vector comprises a genetically engineered form of an adenovirus. Knowledge of the genetic organization of adenovirus, a 36 kb, linear, double-stranded DNA virus, allows substitution of large pieces of adenoviral DNA with foreign sequences up to 7 kb (Grunhaus & Horwitz, 1992). In contrast to retrovirus, the adenoviral infection of host cells does not result in chromosomal integration because adenoviral DNA can replicate in an episomal manner without potential genotoxicity. Also, adenoviruses are structurally stable, and no genome rearrangement has been detected after extensive amplification. Adenovirus can infect virtually all epithelial cells regardless of their cell cycle stage. So far, adenoviral infection appears to be linked only to mild disease such as acute respiratory disease in humans.

Adenovirus is particularly suitable for use as a gene transfer vector because of its mid-sized genome, ease of manipulation, high titer, wide target-cell range and high infectivity. Both ends of the viral genome contain 100-200 base pair inverted repeats (ITRs), which are *cis* elements necessary for viral DNA replication and packaging. The early (E) and late (L) regions of the genome contain different transcription units that are divided by the onset of viral DNA replication. The E1 region (E1A and E1B) encodes proteins responsible for the regulation of transcription of the viral genome and a few cellular genes. The expression of the E2 region (E2A and E2B) results in the synthesis of the proteins for viral DNA replication. These proteins are involved in DNA replication, late gene expression and host cell shut-off (Renan, 1990). The products of the late genes, including the majority of the viral capsid proteins, are expressed only after significant processing of a single primary

transcript issued by the major late promoter (MLP). The MLP, (located at 16.8 m.u.) is particularly efficient during the late phase of infection, and all the mRNA's issued from this promoter possess a 5'-tripartite leader (TPL) sequence which makes them preferred mRNA's for translation.

5 In a current system, recombinant adenovirus is generated from homologous recombination between shuttle vector and provirus vector. Due to the possible recombination between two proviral vectors, wild-type adenovirus may be generated from this process. Therefore, it is critical to isolate a single clone of virus from an individual plaque and examine its genomic structure.

10 Generation and propagation of the current adenovirus vectors, which are replication deficient, depend on a unique helper cell line, designated 293, which was transformed from human embryonic kidney cells by Ad5 DNA fragments and constitutively expresses E1 proteins (Graham *et al.*, 1977). Since the E3 region is dispensable from the adenovirus genome (Jones & Shenk, 1978), the current adenovirus vectors, with the help of 15 293 cells, carry foreign DNA in either the E1, the D3 or both regions (Graham & Prevec, 1991). In nature, adenovirus can package approximately 105% of the wild-type genome (Ghosh-Choudhury *et al.*, 1987), providing capacity for about 2 extra kB of DNA. Combined with the approximately 5.5 kB of DNA that is replaceable in the E1 and E3 regions, the maximum capacity of the current adenovirus vector is under 7.5 kB, or about 15% of the total 20 length of the vector. More than 80% of the adenovirus viral genome remains in the vector backbone and is the source of vector-borne cytotoxicity. Also, the replication deficiency of the E1-deleted virus is incomplete. For example, leakage of viral gene expression has been observed with the currently available vectors at high multiplicities of infection (MOI) (Mulligan, 1993).

25 Helper cell lines may be derived from human cells such as human embryonic kidney cells, muscle cells, hematopoietic cells or other human embryonic mesenchymal or epithelial cells. Alternatively, the helper cells may be derived from the cells of other mammalian species that are permissive for human adenovirus. Such cells include, *e.g.*, Vero cells or other monkey embryonic mesenchymal or epithelial cells. As stated above, the 30 currently preferred helper cell line is 293.

Recently, Racher *et al.* (1995) disclosed improved methods for culturing 293 cells and propagating adenovirus. In one format, natural cell aggregates are grown by inoculating individual cells into 1 liter siliconized spinner flasks (Techne, Cambridge, UK)

containing 100-200 ml of medium. Following stirring at 40 rpm, the cell viability is estimated with trypan blue. In another format, Fibra-Cel microcarriers (Bibby Sterlin, Stone, UK) (5 g/l) is employed as follows. A cell inoculum, resuspended in 5 ml of medium, is added to the carrier (50 ml) in a 250 ml Erlenmeyer flask and left stationary, with occasional agitation, for 1 to 4 h. The medium is then replaced with 50 ml of fresh medium and shaking initiated. For virus production, cells are allowed to grow to about 80% confluence, after which time the medium is replaced (to 25% of the final volume) and adenovirus added at an MOI of 0.05. Cultures are left stationary overnight, following which the volume is increased to 100% and shaking commenced for another 72 h.

Other than the requirement that the adenovirus vector be replication defective, or at least conditionally defective, the nature of the adenovirus vector is not believed to be crucial to the successful practice of the invention. The adenovirus may be of any of the 42 different known serotypes or subgroups A-F. Adenovirus type 5 of subgroup C is the preferred starting material in order to obtain a conditional replication-defective adenovirus vector for use in the present invention, since Adenovirus type 5 is a human adenovirus about which a great deal of biochemical and genetic information is known, and it has historically been used for most constructions employing adenovirus as a vector.

As stated above, the typical vector according to the present invention is replication defective and will not have an adenovirus E1 region. Thus, it will be most convenient to introduce the polynucleotide encoding the gene of interest at the position from which the E1-coding sequences have been removed. However, the position of insertion of the construct within the adenovirus sequences is not critical to the invention. The polynucleotide encoding the gene of interest may also be inserted in lieu of the deleted E3 region in E3 replacement vectors as described by Karlsson *et al.* (1986) or in the E4 region where a helper cell line or helper virus complements the E4 defect.

Adenovirus is easy to grow and manipulate and exhibits broad host range *in vitro* and *in vivo*. This group of viruses can be obtained in high titers, e.g., 10^9 - 10^{11} plaque-forming units per ml, and they are highly infective. The life cycle of adenovirus does not require integration into the host cell genome. The foreign genes delivered by adenovirus vectors are episomal and, therefore, have low genotoxicity to host cells. No side effects have been reported in studies of vaccination with wild-type adenovirus (Couch *et al.*, 1963; Top *et al.*, 1971), demonstrating their safety and therapeutic potential as *in vivo* gene transfer vectors.

Adenovirus vectors have been used in eukaryotic gene expression (Levrero *et al.*, 1991; Gomez-Foix *et al.*, 1992) and vaccine development (Grunhaus & Horwitz, 1992; Graham & Prevec, 1992). Recently, animal studies suggested that recombinant adenovirus could be used for gene therapy (Stratford-Perricaudet & Perricaudet, 1991; Stratford-Perricaudet *et al.*, 1990; Rich *et al.*, 1993). Studies in administering recombinant adenovirus to different tissues include trachea instillation (Rosenfeld *et al.*, 1991; Rosenfeld *et al.*, 1992), muscle injection (Ragot *et al.*, 1993), peripheral intravenous injections (Herz & Gerard, 1993) and stereotactic inoculation into the brain (Le Gal La Salle *et al.*, 1993).

B. Retroviruses

The retroviruses are a group of single-stranded RNA viruses characterized by an ability to convert their RNA to double-stranded DNA in infected cells by a process of reverse-transcription (Coffin, 1990). The resulting DNA then stably integrates into cellular chromosomes as a provirus and directs synthesis of viral proteins. The integration results in the retention of the viral gene sequences in the recipient cell and its descendants. The retroviral genome contains three genes, gag, pol, and env that code for capsid proteins, polymerase enzyme, and envelope components, respectively. A sequence found upstream from the gag gene contains a signal for packaging of the genome into virions. Two long terminal repeat (LTR) sequences are present at the 5' and 3' ends of the viral genome. These contain strong promoter and enhancer sequences and are also required for integration in the host cell genome (Coffin, 1990).

In order to construct a retroviral vector, a nucleic acid encoding one or more oligonucleotide or polynucleotide sequences of interest is inserted into the viral genome in the place of certain viral sequences to produce a virus that is replication-defective. In order to produce virions, a packaging cell line containing the gag, pol, and env genes but without the LTR and packaging components is constructed (Mann *et al.*, 1983). When a recombinant plasmid containing a cDNA, together with the retroviral LTR and packaging sequences is introduced into this cell line (by calcium phosphate precipitation for example), the packaging sequence allows the RNA transcript of the recombinant plasmid to be packaged into viral particles, which are then secreted into the culture media (Nicolas & Rubenstein, 1988; Temin, 1986; Mann *et al.*, 1983). The media containing the recombinant retroviruses is then collected, optionally concentrated, and used for gene transfer. Retroviral vectors are able to infect a broad variety of cell types. However, integration and stable expression require the division of host cells (Paskind *et al.*, 1975).

A novel approach designed to allow specific targeting of retrovirus vectors was recently developed based on the chemical modification of a retrovirus by the chemical addition of lactose residues to the viral envelope. This modification could permit the specific infection of hepatocytes *via* sialoglycoprotein receptors.

5 A different approach to targeting of recombinant retroviruses was designed in which biotinylated antibodies against a retroviral envelope protein and against a specific cell receptor were used. The antibodies were coupled *via* the biotin components by using streptavidin (Roux *et al.*, 1989). Using antibodies against major histocompatibility complex class I and class II antigens, they demonstrated the infection of a variety of human cells that
10 bore those surface antigens with an ecotropic virus *in vitro* (Roux *et al.*, 1989).

C. Adeno-Associated Viruses

AAV (Ridgeway, 1988; Hermonat & Muzychka, 1984) is a parovirus, discovered as a contamination of adenoviral stocks. It is a ubiquitous virus (antibodies are present in 85% of the US human population) that has not been linked to any disease. It is
15 also classified as a dependovirus, because its replication is dependent on the presence of a helper virus, such as adenovirus. Five serotypes have been isolated, of which AAV-2 is the best characterized. AAV has a single-stranded linear DNA that is encapsidated into capsid proteins VP1, VP2 and VP3 to form an icosahedral virion of 20 to 24 nm in diameter (Muzychka & McLaughlin, 1988).

20 The AAV DNA is approximately 4.7 kilobases long. It contains two open reading frames and is flanked by two ITRs. There are two major genes in the AAV genome: *rep* and *cap*. The *rep* gene codes for proteins responsible for viral replication, whereas *cap* codes for capsid protein VP1-3. Each ITR forms a T-shaped hairpin structure. These terminal repeats are the only essential *cis* components of the AAV for chromosomal
25 integration. Therefore, the AAV can be used as a vector with all viral coding sequences removed and replaced by the cassette of genes for delivery. Three viral promoters have been identified and named p5, p19, and p40, according to their map position. Transcription from p5 and p19 results in production of *rep* proteins, and transcription from p40 produces the capsid proteins (Hermonat & Muzychka, 1984).

30 There are several factors that prompted researchers to study the possibility of using rAAV as an expression vector. One is that the requirements for delivering a gene to integrate into the host chromosome are surprisingly few. It is necessary to have the 145-bp ITRs, which are only 6% of the AAV genome. This leaves room in the vector to assemble a

4.5-kb DNA insertion. While this carrying capacity may prevent the AAV from delivering large genes, it is amply suited for delivering the antisense constructs of the present invention.

AAV is also a good choice of delivery vehicles due to its safety. There is a relatively complicated rescue mechanism: not only wild type adenovirus but also AAV genes are required to mobilize rAAV. Likewise, AAV is not pathogenic and not associated with any disease. The removal of viral coding sequences minimizes immune reactions to viral gene expression, and therefore, rAAV does not evoke an inflammatory response.

D. Other Viral Vectors as Expression Constructs

Other viral vectors may be employed as expression constructs in the present invention for the delivery of oligonucleotide or polynucleotide sequences to a host cell.

Vectors derived from viruses such as vaccinia virus (Ridgeway, 1988; Coupar *et al.*, 1988), lentiviruses, polio viruses and herpes viruses may be employed. They offer several attractive features for various mammalian cells (Friedmann, 1989; Ridgeway, 1988; Coupar *et al.*, 1988; Horwitz *et al.*, 1990).

With the recent recognition of defective hepatitis B viruses, new insight was gained into the structure-function relationship of different viral sequences. *In vitro* studies showed that the virus could retain the ability for helper-dependent packaging and reverse transcription despite the deletion of up to 80% of its genome (Horwitz *et al.*, 1990). This suggested that large portions of the genome could be replaced with foreign genetic material.

The hepatotropism and persistence (integration) were particularly attractive properties for liver-directed gene transfer. Chang *et al.* (1991) introduced the chloramphenicol acetyltransferase (CAT) gene into duck hepatitis B virus genome in the place of the polymerase, surface, and pre-surface coding sequences. It was cotransfected with wild-type virus into an avian hepatoma cell line. Culture media containing high titers of the recombinant virus were used to infect primary duckling hepatocytes. Stable CAT gene expression was detected for at least 24 days after transfection (Chang *et al.*, 1991).

E. Non-viral vectors

In order to effect expression of the oligonucleotide or polynucleotide sequences of the present invention, the expression construct must be delivered into a cell.

This delivery may be accomplished *in vitro*, as in laboratory procedures for transforming cells lines, or *in vivo* or *ex vivo*, as in the treatment of certain disease states. As described above, one preferred mechanism for delivery is *via* viral infection where the expression construct is encapsulated in an infectious viral particle.

Once the expression construct has been delivered into the cell the nucleic acid encoding the desired oligonucleotide or polynucleotide sequences may be positioned and expressed at different sites. In certain embodiments, the nucleic acid encoding the construct may be stably integrated into the genome of the cell. This integration may be in the specific location and orientation *via* homologous recombination (gene replacement) or it may be integrated in a random, non-specific location (gene augmentation). In yet further embodiments, the nucleic acid may be stably maintained in the cell as a separate, episomal segment of DNA. Such nucleic acid segments or “episomes” encode sequences sufficient to permit maintenance and replication independent of or in synchronization with the host cell cycle. How the expression construct is delivered to a cell and where in the cell the nucleic acid remains is dependent on the type of expression construct employed.

In certain embodiments of the invention, the expression construct comprising one or more oligonucleotide or polynucleotide sequences may simply consist of naked recombinant DNA or plasmids. Transfer of the construct may be performed by any of the methods mentioned above which physically or chemically permeabilize the cell membrane. This is particularly applicable for transfer *in vitro* but it may be applied to *in vivo* use as well. Dubensky *et al.* (1984) successfully injected polyomavirus DNA in the form of calcium phosphate precipitates into liver and spleen of adult and newborn mice demonstrating active viral replication and acute infection. Benvenisty & Reshef (1986) also demonstrated that direct intraperitoneal injection of calcium phosphate-precipitated plasmids results in expression of the transfected genes. It is envisioned that DNA encoding a gene of interest may also be transferred in a similar manner *in vivo* and express the gene product.

Another embodiment of the invention for transferring a naked DNA expression construct into cells may involve particle bombardment. This method depends on the ability to accelerate DNA-coated microprojectiles to a high velocity allowing them to pierce cell membranes and enter cells without killing them (Klein *et al.*, 1987). Several devices for accelerating small particles have been developed. One such device relies on a high voltage discharge to generate an electrical current, which in turn provides the motive force (Yang *et al.*, 1990). The microprojectiles used have consisted of biologically inert substances such as tungsten or gold beads.

Selected organs including the liver, skin, and muscle tissue of rats and mice have been bombarded *in vivo* (Yang *et al.*, 1990; Zelenin *et al.*, 1991). This may require surgical exposure of the tissue or cells, to eliminate any intervening tissue between the gun

and the target organ, *i.e.*, *ex vivo* treatment. Again, DNA encoding a particular gene may be delivered *via* this method and still be incorporated by the present invention.

VII. POLYPEPTIDE COMPOSITIONS

The present invention, in other aspects, provides polypeptide compositions.

- 5 Generally, a polypeptide of the invention will be an isolated polypeptide (or an epitope, variant, or active fragment thereof) derived from a mammalian species. Preferably, the polypeptide is encoded by a polynucleotide sequence disclosed herein or a sequence which hybridizes under moderately stringent conditions to a polynucleotide sequence disclosed herein. Alternatively, the polypeptide may be defined as a polypeptide which comprises a
10 contiguous amino acid sequence from an amino acid sequence disclosed herein, or which polypeptide comprises an entire amino acid sequence disclosed herein.

Immunogenic portions may generally be identified using well known techniques, such as those summarized in Paul, *Fundamental Immunology*, 3rd ed., 243-247 (1993) and references cited therein. Such techniques include screening polypeptides for the ability to react with antigen-specific antibodies, antisera and/or T-cell lines or clones. As used herein, antisera and antibodies are “antigen-specific” if they specifically bind to an antigen (*i.e.*, they react with the protein in an ELISA or other immunoassay, and do not react detectably with unrelated proteins). Such antisera and antibodies may be prepared as described herein, and using well known techniques. An immunogenic portion of a
15 *Mycobacterium* sp. protein is a portion that reacts with such antisera and/or T-cells at a level that is not substantially less than the reactivity of the full length polypeptide (*e.g.*, in an ELISA and/or T-cell reactivity assay). Such immunogenic portions may react within such assays at a level that is similar to or greater than the reactivity of the full length polypeptide. Such screens may generally be performed using methods well known to those of ordinary
20 skill in the art, such as those described in Harlow & Lane, *Antibodies: A Laboratory Manual* (1988). For example, a polypeptide may be immobilized on a solid support and contacted with patient sera to allow binding of antibodies within the sera to the immobilized polypeptide. Unbound sera may then be removed and bound antibodies detected using, for example, ¹²⁵I-labeled Protein A.
25 Polypeptides may be prepared using any of a variety of well known techniques. Recombinant polypeptides encoded by DNA sequences as described above may be readily prepared from the DNA sequences using any of a variety of expression vectors known to those of ordinary skill in the art. Expression may be achieved in any appropriate

host cell that has been transformed or transfected with an expression vector containing a DNA molecule that encodes a recombinant polypeptide. Suitable host cells include prokaryotes, yeast, and higher eukaryotic cells, such as mammalian cells and plant cells. Preferably, the host cells employed are *E. coli*, yeast or a mammalian cell line such as COS or CHO. Supernatants from suitable host/vector systems which secrete recombinant protein or polypeptide into culture media may be first concentrated using a commercially available filter. Following concentration, the concentrate may be applied to a suitable purification matrix such as an affinity matrix or an ion exchange resin. Finally, one or more reverse phase HPLC steps can be employed to further purify a recombinant polypeptide.

10 Polypeptides of the invention, immunogenic fragments thereof, and other variants having less than about 100 amino acids, and generally less than about 50 amino acids, may also be generated by synthetic means, using techniques well known to those of ordinary skill in the art. For example, such polypeptides may be synthesized using any of the commercially available solid-phase techniques, such as the Merrifield solid-phase synthesis method, where amino acids are sequentially added to a growing amino acid chain. *See* Merrifield, *J. Am. Chem. Soc.* 85:2149-2146 (1963). Equipment for automated synthesis of polypeptides is commercially available from suppliers such as Perkin Elmer/Applied BioSystems Division (Foster City, CA), and may be operated according to the manufacturer's instructions.

15 20 Within certain specific embodiments, a polypeptide may be a fusion protein that comprises multiple polypeptides as described herein, or that comprises at least one polypeptide as described herein and an unrelated sequence, such as a known tumor protein. A fusion partner may, for example, assist in providing T helper epitopes (an immunological fusion partner), preferably T helper epitopes recognized by humans, or may assist in expressing the protein (an expression enhancer) at higher yields than the native recombinant protein. Certain preferred fusion partners are both immunological and expression enhancing fusion partners. Other fusion partners may be selected so as to increase the solubility of the protein or to enable the protein to be targeted to desired intracellular compartments. Still further fusion partners include affinity tags, which facilitate purification of the protein.

25 30 Fusion proteins may generally be prepared using standard techniques, including chemical conjugation. Preferably, a fusion protein is expressed as a recombinant protein, allowing the production of increased levels, relative to a non-fused protein, in an expression system. Briefly, DNA sequences encoding the polypeptide components may be assembled separately, and ligated into an appropriate expression vector. The 3' end of the

DNA sequence encoding one polypeptide component is ligated, with or without a peptide linker, to the 5' end of a DNA sequence encoding the second polypeptide component so that the reading frames of the sequences are in phase. This permits translation into a single fusion protein that retains the biological activity of both component polypeptides.

5 A peptide linker sequence may be employed to separate the first and second polypeptide components by a distance sufficient to ensure that each polypeptide folds into its secondary and tertiary structures. Such a peptide linker sequence is incorporated into the fusion protein using standard techniques well known in the art. Suitable peptide linker sequences may be chosen based on the following factors: (1) their ability to adopt a flexible
10 extended conformation; (2) their inability to adopt a secondary structure that could interact with functional epitopes on the first and second polypeptides; and (3) the lack of hydrophobic or charged residues that might react with the polypeptide functional epitopes. Preferred peptide linker sequences contain Gly, Asn and Ser residues. Other near neutral amino acids, such as Thr and Ala may also be used in the linker sequence. Amino acid sequences which
15 may be usefully employed as linkers include those disclosed in Maratea *et al.*, *Gene* 40:39-46 (1985); Murphy *et al.*, *Proc. Natl. Acad. Sci. USA* 83:8258-8262 (1986); U.S. Patent No. 4,935,233 and U.S. Patent No. 4,751,180. The linker sequence may generally be from 1 to about 50 amino acids in length. Linker sequences are not required when the first and second polypeptides have non-essential N-terminal amino acid regions that can be used to separate
20 the functional domains and prevent steric interference.

The ligated DNA sequences are operably linked to suitable transcriptional or translational regulatory elements. The regulatory elements responsible for expression of DNA are located only 5' to the DNA sequence encoding the first polypeptides. Similarly, stop codons required to end translation and transcription termination signals are only present
25 3' to the DNA sequence encoding the second polypeptide.

Fusion proteins are also provided. Such proteins comprise a polypeptide as described herein together with an unrelated immunogenic protein. Preferably the immunogenic protein is capable of eliciting a recall response. Examples of such proteins include tetanus, tuberculosis and hepatitis proteins (*see, e.g.*, Stoute *et al.*, *New Engl. J. Med.* 336:86-91 (1997)).

Within preferred embodiments, an immunological fusion partner is derived from protein D, a surface protein of the gram-negative bacterium *Haemophilus influenza* B (WO 91/18926). Preferably, a protein D derivative comprises approximately the first third of the protein (*e.g.*, the first N-terminal 100-110 amino acids), and a protein D derivative may

be lipidated. Within certain preferred embodiments, the first 109 residues of a lipoprotein D fusion partner is included on the N-terminus to provide the polypeptide with additional exogenous T-cell epitopes and to increase the expression level in *E. coli* (thus functioning as an expression enhancer). The lipid tail ensures optimal presentation of the antigen to antigen presenting cells. Other fusion partners include the non-structural protein from influenzae virus, NS1 (hemagglutinin). Typically, the N-terminal 81 amino acids are used, although different fragments that include T-helper epitopes may be used.

In another embodiment, the immunological fusion partner is the protein known as LYTA, or a portion thereof (preferably a C-terminal portion). LYTA is derived from *Streptococcus pneumoniae*, which synthesizes an N-acetyl-L-alanine amidase known as amidase LYTA (encoded by the LytA gene; *Gene* 43:265-292 (1986)). LYTA is an autolysin that specifically degrades certain bonds in the peptidoglycan backbone. The C-terminal domain of the LYTA protein is responsible for the affinity to the choline or to some choline analogues such as DEAE. This property has been exploited for the development of *E. coli* C-LYTA expressing plasmids useful for expression of fusion proteins. Purification of hybrid proteins containing the C-LYTA fragment at the amino terminus has been described (*see Biotechnology* 10:795-798 (1992)). Within a preferred embodiment, a repeat portion of LYTA may be incorporated into a fusion protein. A repeat portion is found in the C-terminal region starting at residue 178. A particularly preferred repeat portion incorporates residues 188-305.

In general, polypeptides (including fusion proteins) and polynucleotides as described herein are isolated. An "isolated" polypeptide or polynucleotide is one that is removed from its original environment. For example, a naturally-occurring protein is isolated if it is separated from some or all of the coexisting materials in the natural system. Preferably, such polypeptides are at least about 90% pure, more preferably at least about 95% pure and most preferably at least about 99% pure. A polynucleotide is considered to be isolated if, for example, it is cloned into a vector that is not a part of the natural environment.

VIII. T CELLS

Immunotherapeutic compositions may also, or alternatively, comprise T cells specific for a *Mycobacterium* antigen. Such cells may generally be prepared *in vitro* or *ex vivo*, using standard procedures. For example, T cells may be isolated from bone marrow, peripheral blood, or a fraction of bone marrow or peripheral blood of a patient, using a commercially available cell separation system, such as the Isolex™ System, available from

Nexell Therapeutics, Inc. (Irvine, CA; *see also* U.S. Patent No. 5,240,856; U.S. Patent No. 5,215,926; WO 89/06280; WO 91/16116 and WO 92/07243). Alternatively, T cells may be derived from related or unrelated humans, non-human mammals, cell lines or cultures.

T cells may be stimulated with a polypeptide of the invention, polynucleotide 5 encoding such a polypeptide, and/or an antigen presenting cell (APC) that expresses such a polypeptide. Such stimulation is performed under conditions and for a time sufficient to permit the generation of T cells that are specific for the polypeptide. Preferably, the polypeptide or polynucleotide is present within a delivery vehicle, such as a microsphere, to facilitate the generation of specific T cells.

10 T cells are considered to be specific for a polypeptide of the invention if the T cells specifically proliferate, secrete cytokines or kill target cells coated with the polypeptide or expressing a gene encoding the polypeptide. T cell specificity may be evaluated using any of a variety of standard techniques. For example, within a chromium release assay or proliferation assay, a stimulation index of more than two fold increase in lysis and/or 15 proliferation, compared to negative controls, indicates T cell specificity. Such assays may be performed, for example, as described in Chen *et al.*, *Cancer Res.* 54:1065-1070 (1994). Alternatively, detection of the proliferation of T cells may be accomplished by a variety of known techniques. For example, T cell proliferation can be detected by measuring an increased rate of DNA synthesis (*e.g.*, by pulse-labeling cultures of T cells with tritiated 20 thymidine and measuring the amount of tritiated thymidine incorporated into DNA). Contact with a polypeptide of the invention (100 ng/ml - 100 µg/ml, preferably 200 ng/ml-25 µg/ml) for 3 - 7 days should result in at least a two fold increase in proliferation of the T cells. Contact as described above for 2-3 hours should result in activation of the T cells, as measured using standard cytokine assays in which a two fold increase in the level of cytokine 25 release (*e.g.*, TNF or IFN- γ) is indicative of T cell activation (*see* Coligan *et al.*, *Current Protocols in Immunology*, vol. 1 (1998)). T cells that have been activated in response to a polypeptide, polynucleotide or polypeptide-expressing APC may be CD4 $^{+}$ and/or CD8 $^{+}$. Protein-specific T cells may be expanded using standard techniques. Within preferred 30 embodiments, the T cells are derived from a patient, a related donor or an unrelated donor, and are administered to the patient following stimulation and expansion.

For therapeutic purposes, CD4 $^{+}$ or CD8 $^{+}$ T cells that proliferate in response to a polypeptide, polynucleotide or APC can be expanded in number either *in vitro* or *in vivo*. Proliferation of such T cells *in vitro* may be accomplished in a variety of ways. For example,

RECEIVED
U.S. PATENT AND TRADEMARK OFFICE

the T cells can be re-exposed to a polypeptide, or a short peptide corresponding to an immunogenic portion of such a polypeptide, with or without the addition of T cell growth factors, such as interleukin-2, and/or stimulator cells that synthesize a r polypeptide. Alternatively, one or more T cells that proliferate in the presence of ar protein can be 5 expanded in number by cloning. Methods for cloning cells are well known in the art, and include limiting dilution.

IX. PHARMACEUTICAL COMPOSITIONS

In additional embodiments, the present invention concerns formulation of one or more of the polynucleotide, polypeptide, T-cell and/or antibody compositions disclosed 10 herein in pharmaceutically-acceptable solutions for administration to a cell or an animal, either alone, or in combination with one or more other modalities of therapy.

It will also be understood that, if desired, the nucleic acid segment, RNA, DNA or PNA compositions that express a polypeptide as disclosed herein may be administered in combination with other agents as well, such as, e.g., other proteins or 15 polypeptides or various pharmaceutically-active agents. In fact, there is virtually no limit to other components that may also be included, given that the additional agents do not cause a significant adverse effect upon contact with the target cells or host tissues. The compositions may thus be delivered along with various other agents as required in the particular instance. Such compositions may be purified from host cells or other biological sources, or 20 alternatively may be chemically synthesized as described herein. Likewise, such compositions may further comprise substituted or derivatized RNA or DNA compositions.

Formulation of pharmaceutically-acceptable excipients and carrier solutions is well-known to those of skill in the art, as is the development of suitable dosing and treatment regimens for using the particular compositions described herein in a variety of treatment 25 regimens, including e.g., oral, parenteral, intravenous, intranasal, and intramuscular administration and formulation.

A. Oral Delivery

In certain applications, the pharmaceutical compositions disclosed herein may be delivered *via* oral administration to an animal. As such, these compositions may be 30 formulated with an inert diluent or with an assimilable edible carrier, or they may be enclosed in hard- or soft-shell gelatin capsule, or they may be compressed into tablets, or they may be incorporated directly with the food of the diet.

5
10
15

The active compounds may even be incorporated with excipients and used in the form of ingestible tablets, buccal tables, troches, capsules, elixirs, suspensions, syrups, wafers, and the like (Mathiowitz *et al.*, 1997; Hwang *et al.*, 1998; U. S. Patent 5,641,515; U. S. Patent 5,580,579 and U. S. Patent 5,792,451, each specifically incorporated herein by reference in its entirety). The tablets, troches, pills, capsules and the like may also contain the following: a binder, as gum tragacanth, acacia, cornstarch, or gelatin; excipients, such as dicalcium phosphate; a disintegrating agent, such as corn starch, potato starch, alginic acid and the like; a lubricant, such as magnesium stearate; and a sweetening agent, such as sucrose, lactose or saccharin may be added or a flavoring agent, such as peppermint, oil of wintergreen, or cherry flavoring. When the dosage unit form is a capsule, it may contain, in addition to materials of the above type, a liquid carrier. Various other materials may be present as coatings or to otherwise modify the physical form of the dosage unit. For instance, tablets, pills, or capsules may be coated with shellac, sugar, or both. A syrup of elixir may contain the active compound sucrose as a sweetening agent methyl and propylparabens as preservatives, a dye and flavoring, such as cherry or orange flavor. Of course, any material used in preparing any dosage unit form should be pharmaceutically pure and substantially non-toxic in the amounts employed. In addition, the active compounds may be incorporated into sustained-release preparation and formulations.

20
25

Typically, these formulations may contain at least about 0.1% of the active compound or more, although the percentage of the active ingredient(s) may, of course, be varied and may conveniently be between about 1 or 2% and about 60% or 70% or more of the weight or volume of the total formulation. Naturally, the amount of active compound(s) in each therapeutically useful composition may be prepared in such a way that a suitable dosage will be obtained in any given unit dose of the compound. Factors such as solubility, bioavailability, biological half-life, route of administration, product shelf life, as well as other pharmacological considerations will be contemplated by one skilled in the art of preparing such pharmaceutical formulations, and as such, a variety of dosages and treatment regimens may be desirable.

30

For oral administration the compositions of the present invention may alternatively be incorporated with one or more excipients in the form of a mouthwash, dentifrice, buccal tablet, oral spray, or sublingual orally-administered formulation. For example, a mouthwash may be prepared incorporating the active ingredient in the required amount in an appropriate solvent, such as a sodium borate solution (Dobell's Solution). Alternatively, the active ingredient may be incorporated into an oral solution such as one

containing sodium borate, glycerin and potassium bicarbonate, or dispersed in a dentifrice, or added in a therapeutically-effective amount to a composition that may include water, binders, abrasives, flavoring agents, foaming agents, and humectants. Alternatively the compositions may be fashioned into a tablet or solution form that may be placed under the tongue or

5 otherwise dissolved in the mouth.

B. Injectable Delivery

In certain circumstances it will be desirable to deliver the pharmaceutical

compositions disclosed herein parenterally, intravenously, intramuscularly, or even

intraperitoneally as described in U. S. Patent 5,543,158; U. S. Patent 5,641,515 and U. S.

10 Patent 5,399,363 (each specifically incorporated herein by reference in its entirety).

Solutions of the active compounds as free base or pharmacologically acceptable salts may be prepared in water suitably mixed with a surfactant, such as hydroxypropylcellulose.

Dispersions may also be prepared in glycerol, liquid polyethylene glycols, and mixtures thereof and in oils. Under ordinary conditions of storage and use, these preparations contain

15 a preservative to prevent the growth of microorganisms.

The pharmaceutical forms suitable for injectable use include sterile aqueous solutions or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions (U. S. Patent 5,466,468, specifically incorporated herein by reference in its entirety). In all cases the form must be sterile and must be fluid to the extent

20 that easy syringability exists. It must be stable under the conditions of manufacture and

storage and must be preserved against the contaminating action of microorganisms, such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (*e.g.*, glycerol, propylene glycol, and liquid polyethylene

glycol, and the like), suitable mixtures thereof, and/or vegetable oils. Proper fluidity may be

25 maintained, for example, by the use of a coating, such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. The prevention of the action of microorganisms can be facilitated by various antibacterial and antifungal

agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium

30 chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

P A T E N T

For parenteral administration in an aqueous solution, for example, the solution should be suitably buffered if necessary and the liquid diluent first rendered isotonic with sufficient saline or glucose. These particular aqueous solutions are especially suitable for intravenous, intramuscular, subcutaneous and intraperitoneal administration. In this connection, a sterile aqueous medium that can be employed will be known to those of skill in the art in light of the present disclosure. For example, one dosage may be dissolved in 1 ml of isotonic NaCl solution and either added to 1000 ml of hypodermoclysis fluid or injected at the proposed site of infusion (*see, e.g., Remington's Pharmaceutical Sciences*, 15th Edition, pp. 1035-1038 and 1570-1580). Some variation in dosage will necessarily occur depending on the condition of the subject being treated. The person responsible for administration will, in any event, determine the appropriate dose for the individual subject. Moreover, for human administration, preparations should meet sterility, pyrogenicity, and the general safety and purity standards as required by FDA Office of Biologics standards.

Sterile injectable solutions are prepared by incorporating the active compounds in the required amount in the appropriate solvent with various of the other ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the various sterilized active ingredients into a sterile vehicle which contains the basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum-drying and freeze-drying techniques which yield a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

The compositions disclosed herein may be formulated in a neutral or salt form. Pharmaceutically-acceptable salts, include the acid addition salts (formed with the free amino groups of the protein) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, histidine, procaine and the like. Upon formulation, solutions will be administered in a manner compatible with the dosage formulation and in such amount as is therapeutically effective. The formulations are easily administered in a variety of dosage forms such as injectable solutions, drug-release capsules, and the like.

As used herein, “carrier” includes any and all solvents, dispersion media, vehicles, coatings, diluents, antibacterial and antifungal agents, isotonic and absorption delaying agents, buffers, carrier solutions, suspensions, colloids, and the like. The use of such media and agents for pharmaceutical active substances is well known in the art. Except 5 insofar as any conventional media or agent is incompatible with the active ingredient, its use in the therapeutic compositions is contemplated. Supplementary active ingredients can also be incorporated into the compositions.

The phrase “pharmaceutically-acceptable” refers to molecular entities and compositions that do not produce an allergic or similar untoward reaction when administered 10 to a human. The preparation of an aqueous composition that contains a protein as an active ingredient is well understood in the art. Typically, such compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid prior to injection can also be prepared. The preparation can also be emulsified.

15 **C. Nasal Delivery**

In certain embodiments, the pharmaceutical compositions may be delivered by intranasal sprays, inhalation, and/or other aerosol delivery vehicles. Methods for delivering genes, nucleic acids, and peptide compositions directly to the lungs *via* nasal aerosol sprays has been described *e.g.*, in U. S. Patent 5,756,353 and U. S. Patent 5,804,212 (each 20 specifically incorporated herein by reference in its entirety). Likewise, the delivery of drugs using intranasal microparticle resins (Takenaga *et al.*, 1998) and lysophosphatidyl-glycerol compounds (U. S. Patent 5,725,871, specifically incorporated herein by reference in its entirety) are also well-known in the pharmaceutical arts. Likewise, transmucosal drug delivery in the form of a polytetrafluoroethylene support matrix is described in U. S. Patent 25 5,780,045 (specifically incorporated herein by reference in its entirety).

D. Liposome-, Nanocapsule-, and Microparticle-Mediated Delivery

In certain embodiments, the inventors contemplate the use of liposomes, nanocapsules, microparticles, microspheres, lipid particles, vesicles, and the like, for the introduction of the compositions of the present invention into suitable host cells. In 30 particular, the compositions of the present invention may be formulated for delivery either encapsulated in a lipid particle, a liposome, a vesicle, a nanosphere, or a nanoparticle or the like.

Such formulations may be preferred for the introduction of pharmaceutically-acceptable formulations of the nucleic acids or constructs disclosed herein. The formation and use of liposomes is generally known to those of skill in the art (see for example, Couvreur *et al.*, 1977; Couvreur, 1988; Lasic, 1998; which describes the use of liposomes and nanocapsules in the targeted antibiotic therapy for intracellular bacterial infections and diseases). Recently, liposomes were developed with improved serum stability and circulation half-times (Gabizon & Papahadjopoulos, 1988; Allen and Choun, 1987; U. S. Patent 5,741,516, specifically incorporated herein by reference in its entirety). Further, various methods of liposome and liposome like preparations as potential drug carriers have been reviewed (Takakura, 1998; Chandran *et al.*, 1997; Margalit, 1995; U. S. Patent 5,567,434; U. S. Patent 5,552,157; U. S. Patent 5,565,213; U. S. Patent 5,738,868 and U. S. Patent 5,795,587, each specifically incorporated herein by reference in its entirety).

Liposomes have been used successfully with a number of cell types that are normally resistant to transfection by other procedures including T cell suspensions, primary hepatocyte cultures and PC 12 cells (Renneisen *et al.*, 1990; Muller *et al.*, 1990). In addition, liposomes are free of the DNA length constraints that are typical of viral-based delivery systems. Liposomes have been used effectively to introduce genes, drugs (Heath & Martin, 1986; Heath *et al.*, 1986; Balazssovis *et al.*, 1989; Fresta & Puglisi, 1996), radiotherapeutic agents (Pikul *et al.*, 1987), enzymes (Imaizumi *et al.*, 1990a; Imaizumi *et al.*, 1990b), viruses (Faller & Baltimore, 1984), transcription factors and allosteric effectors (Nicolau & Gersonde, 1979) into a variety of cultured cell lines and animals. In addition, several successful clinical trials examining the effectiveness of liposome-mediated drug delivery have been completed (Lopez-Berestein *et al.*, 1985a; 1985b; Coune, 1988; Sculier *et al.*, 1988). Furthermore, several studies suggest that the use of liposomes is not associated with autoimmune responses, toxicity or gonadal localization after systemic delivery (Mori & Fukatsu, 1992).

Liposomes are formed from phospholipids that are dispersed in an aqueous medium and spontaneously form multilamellar concentric bilayer vesicles (also termed multilamellar vesicles (MLVs). MLVs generally have diameters of from 25 nm to 4 μ m. Sonication of MLVs results in the formation of small unilamellar vesicles (SUVs) with diameters in the range of 200 to 500 Å, containing an aqueous solution in the core.

Liposomes bear resemblance to cellular membranes and are contemplated for use in connection with the present invention as carriers for the peptide compositions. They

are widely suitable as both water- and lipid-soluble substances can be entrapped, *i.e.* in the aqueous spaces and within the bilayer itself, respectively. It is possible that the drug-bearing liposomes may even be employed for site-specific delivery of active agents by selectively modifying the liposomal formulation.

5 In addition to the teachings of Couvreur *et al.* (1977; 1988), the following information may be utilized in generating liposomal formulations. Phospholipids can form a variety of structures other than liposomes when dispersed in water, depending on the molar ratio of lipid to water. At low ratios the liposome is the preferred structure. The physical characteristics of liposomes depend on pH, ionic strength and the presence of divalent 10 cations. Liposomes can show low permeability to ionic and polar substances, but at elevated temperatures undergo a phase transition which markedly alters their permeability. The phase transition involves a change from a closely packed, ordered structure, known as the gel state, to a loosely packed, less-ordered structure, known as the fluid state. This occurs at a characteristic phase-transition temperature and results in an increase in permeability to ions, 15 sugars and drugs.

 In addition to temperature, exposure to proteins can alter the permeability of liposomes. Certain soluble proteins, such as cytochrome c, bind, deform and penetrate the bilayer, thereby causing changes in permeability. Cholesterol inhibits this penetration of proteins, apparently by packing the phospholipids more tightly. It is contemplated that the 20 most useful liposome formations for antibiotic and inhibitor delivery will contain cholesterol.

 The ability to trap solutes varies between different types of liposomes. For example, MLVs are moderately efficient at trapping solutes, but SUVs are extremely inefficient. SUVs offer the advantage of homogeneity and reproducibility in size distribution, however, and a compromise between size and trapping efficiency is offered by large 25 unilamellar vesicles (LUVs). These are prepared by ether evaporation and are three to four times more efficient at solute entrapment than MLVs.

 In addition to liposome characteristics, an important determinant in entrapping compounds is the physicochemical properties of the compound itself. Polar compounds are trapped in the aqueous spaces and nonpolar compounds bind to the lipid bilayer of the 30 vesicle. Polar compounds are released through permeation or when the bilayer is broken, but nonpolar compounds remain affiliated with the bilayer unless it is disrupted by temperature or exposure to lipoproteins. Both types show maximum efflux rates at the phase transition temperature.

Liposomes interact with cells *via* four different mechanisms: endocytosis by phagocytic cells of the reticuloendothelial system such as macrophages and neutrophils; adsorption to the cell surface, either by nonspecific weak hydrophobic or electrostatic forces, or by specific interactions with cell-surface components; fusion with the plasma cell

5 membrane by insertion of the lipid bilayer of the liposome into the plasma membrane, with simultaneous release of liposomal contents into the cytoplasm; and by transfer of liposomal lipids to cellular or subcellular membranes, or vice versa, without any association of the liposome contents. It often is difficult to determine which mechanism is operative and more than one may operate at the same time.

10 The fate and disposition of intravenously injected liposomes depend on their physical properties, such as size, fluidity, and surface charge. They may persist in tissues for h or days, depending on their composition, and half lives in the blood range from min to several h. Larger liposomes, such as MLVs and LUVs, are taken up rapidly by phagocytic cells of the reticuloendothelial system, but physiology of the circulatory system restrains the 15 exit of such large species at most sites. They can exit only in places where large openings or pores exist in the capillary endothelium, such as the sinusoids of the liver or spleen. Thus, these organs are the predominate site of uptake. On the other hand, SUVs show a broader tissue distribution but still are sequestered highly in the liver and spleen. In general, this *in vivo* behavior limits the potential targeting of liposomes to only those organs and tissues 20 accessible to their large size. These include the blood, liver, spleen, bone marrow, and lymphoid organs.

Targeting is generally not a limitation in terms of the present invention. However, should specific targeting be desired, methods are available for this to be accomplished. Antibodies may be used to bind to the liposome surface and to direct the 25 antibody and its drug contents to specific antigenic receptors located on a particular cell-type surface. Carbohydrate determinants (glycoprotein or glycolipid cell-surface components that play a role in cell-cell recognition, interaction and adhesion) may also be used as recognition sites as they have potential in directing liposomes to particular cell types. Mostly, it is contemplated that intravenous injection of liposomal preparations would be used, but other 30 routes of administration are also conceivable.

Alternatively, the invention provides for pharmaceutically-acceptable nanocapsule formulations of the compositions of the present invention. Nanocapsules can generally entrap compounds in a stable and reproducible way (Henry-Michel and *et al.*, 1987; Quintanar-Guerrero *et al.*, 1998; Douglas *et al.*, 1987). To avoid side effects due to

intracellular polymeric overloading, such ultrafine particles (sized around 0.1 μm) should be designed using polymers able to be degraded *in vivo*. Biodegradable polyalkyl-cyanoacrylate nanoparticles that meet these requirements are contemplated for use in the present invention. Such particles may be easily made, as described (Couvreur *et al.*, 1980; 1988; zur Muhlen 5 *et al.*, 1998; Zambaux *et al.* 1998; Pinto-Alphandry *et al.*, 1995 and U. S. Patent 5,145,684, specifically incorporated herein by reference in its entirety).

X. VACCINES

In certain preferred embodiments of the present invention, vaccines are provided. The vaccines will generally comprise one or more pharmaceutical compositions, 10 such as those discussed above, in combination with an immunostimulant. An immunostimulant may be any substance that enhances or potentiates an immune response (antibody and/or cell-mediated) to an exogenous antigen. Examples of immunostimulants include adjuvants, biodegradable microspheres (*e.g.*, polylactic galactide) and liposomes (into which the compound is incorporated; *see, e.g.*, Fullerton, U.S. Patent No. 4,235,877).

15 Vaccine preparation is generally described in, for example, Powell & Newman, eds., *Vaccine Design* (the subunit and adjuvant approach) (1995). Pharmaceutical compositions and vaccines within the scope of the present invention may also contain other compounds, which may be biologically active or inactive. For example, one or more immunogenic portions of other tumor antigens may be present, either incorporated into a fusion polypeptide or as a 20 separate compound, within the composition or vaccine.

Illustrative vaccines may contain DNA encoding one or more of the polypeptides as described above, such that the polypeptide is generated *in situ*. As noted above, the DNA may be present within any of a variety of delivery systems known to those of ordinary skill in the art, including nucleic acid expression systems, bacteria and viral 25 expression systems. Numerous gene delivery techniques are well known in the art, such as those described by Rolland, *Crit. Rev. Therap. Drug Carrier Systems* 15:143-198 (1998), and references cited therein. Appropriate nucleic acid expression systems contain the necessary DNA sequences for expression in the patient (such as a suitable promoter and terminating signal). Bacterial delivery systems involve the administration of a bacterium (such as 30 *Bacillus-Calmette-Guerrin*) that expresses an immunogenic portion of the polypeptide on its cell surface or secretes such an epitope. In a preferred embodiment, the DNA may be introduced using a viral expression system (*e.g.*, vaccinia or other pox virus, retrovirus, or adenovirus), which may involve the use of a non-pathogenic (defective), replication

competent virus. Suitable systems are disclosed, for example, in Fisher-Hoch *et al.*, *Proc. Natl. Acad. Sci. USA* 86:317-321 (1989); Flexner *et al.*, *Ann. N.Y. Acad. Sci.* 569:86-103 (1989); Flexner *et al.*, *Vaccine* 8:17-21 (1990); U.S. Patent Nos. 4,603,112, 4,769,330, and 5,017,487; WO 89/01973; U.S. Patent No. 4,777,127; GB 2,200,651; EP 0,345,242; WO 5 91/02805; Berkner, *Biotechniques* 6:616-627 (1988); Rosenfeld *et al.*, *Science* 252:431-434 (1991); Kolls *et al.*, *Proc. Natl. Acad. Sci. USA* 91:215-219 (1994); Kass-Eisler *et al.*, *Proc. Natl. Acad. Sci. USA* 90:11498-11502 (1993); Guzman *et al.*, *Circulation* 88:2838-2848 (1993); and Guzman *et al.*, *Cir. Res.* 73:1202-1207 (1993). Techniques for incorporating DNA into such expression systems are well known to those of ordinary skill in the art. The 10 DNA may also be "naked," as described, for example, in Ulmer *et al.*, *Science* 259:1745-1749 (1993) and reviewed by Cohen, *Science* 259:1691-1692 (1993). The uptake of naked DNA may be increased by coating the DNA onto biodegradable beads, which are efficiently transported into the cells. It will be apparent that a vaccine may comprise both a 15 polynucleotide and a polypeptide component. Such vaccines may provide for an enhanced immune response.

It will be apparent that a vaccine may contain pharmaceutically acceptable salts of the polynucleotides and polypeptides provided herein. Such salts may be prepared from pharmaceutically acceptable non-toxic bases, including organic bases (e.g., salts of primary, secondary and tertiary amines and basic amino acids) and inorganic bases (e.g., 20 sodium, potassium, lithium, ammonium, calcium and magnesium salts).

While any suitable carrier known to those of ordinary skill in the art may be employed in the vaccine compositions of this invention, the type of carrier will vary depending on the mode of administration. Compositions of the present invention may be formulated for any appropriate manner of administration, including for example, topical, oral, 25 nasal, intravenous, intracranial, intraperitoneal, subcutaneous or intramuscular administration. For parenteral administration, such as subcutaneous injection, the carrier preferably comprises water, saline, alcohol, a fat, a wax or a buffer. For oral administration, any of the above carriers or a solid carrier, such as mannitol, lactose, starch, magnesium stearate, sodium saccharine, talcum, cellulose, glucose, sucrose, and magnesium carbonate, may be 30 employed. Biodegradable microspheres (e.g., polylactate polyglycolate) may also be employed as carriers for the pharmaceutical compositions of this invention. Suitable biodegradable microspheres are disclosed, for example, in U.S. Patent Nos. 4,897,268; 5,075,109; 5,928,647; 5,811,128; 5,820,883; 5,853,763; 5,814,344 and 5,942,252. One may also employ a carrier comprising the particulate-protein complexes described in U.S. Patent

No. 5,928,647, which are capable of inducing a class I-restricted cytotoxic T lymphocyte responses in a host.

Such compositions may also comprise buffers (e.g., neutral buffered saline or phosphate buffered saline), carbohydrates (e.g., glucose, mannose, sucrose or dextrans),

5 manitol, proteins, polypeptides or amino acids such as glycine, antioxidants, bacteriostats, chelating agents such as EDTA or glutathione, adjuvants (e.g., aluminum hydroxide), solutes that render the formulation isotonic, hypotonic or weakly hypertonic with the blood of a recipient, suspending agents, thickening agents and/or preservatives. Alternatively, compositions of the present invention may be formulated as a lyophilizate. Compounds may
10 also be encapsulated within liposomes using well known technology.

Any of a variety of immunostimulants may be employed in the vaccines of this invention. For example, an adjuvant may be included. Most adjuvants contain a substance designed to protect the antigen from rapid catabolism, such as aluminum hydroxide or mineral oil, and a stimulator of immune responses, such as lipid A, *Bordetella pertussis* or
15 *Mycobacterium* species or *Mycobacterium* derived proteins. For example, delipidated, deglycolipidated *M. vaccae* ("pVac") can be used. In another embodiment, BCG is used. In addition, the vaccine can be administered to a subject previously exposed to BCG. Suitable adjuvants are commercially available as, for example, Freund's Incomplete Adjuvant and Complete Adjuvant (Difco Laboratories, Detroit, MI); Merck Adjuvant 65 (Merck and
20 Company, Inc., Rahway, NJ); AS-2 and derivatives thereof (SmithKline Beecham, Philadelphia, PA); CWS, TDM, Leif, aluminum salts such as aluminum hydroxide gel (alum) or aluminum phosphate; salts of calcium, iron or zinc; an insoluble suspension of acylated tyrosine; acylated sugars; cationically or anionically derivatized polysaccharides; polyphosphazenes; biodegradable microspheres; monophosphoryl lipid A and quill A.
25 Cytokines, such as GM-CSF or interleukin-2, -7, or -12, may also be used as adjuvants.

Within the vaccines provided herein, the adjuvant composition is preferably designed to induce an immune response predominantly of the Th1 type. High levels of Th1-type cytokines (e.g., IFN- γ , TNF α , IL-2 and IL-12) tend to favor the induction of cell mediated immune responses to an administered antigen. In contrast, high levels of Th2-type cytokines (e.g., IL-4, IL-5, IL-6 and IL-10) tend to favor the induction of humoral immune responses. Following application of a vaccine as provided herein, a patient will support an immune response that includes Th1- and Th2-type responses. Within a preferred embodiment, in which a response is predominantly Th1-type, the level of Th1-type cytokines

DRAFT - NOT FOR FILING

will increase to a greater extent than the level of Th2-type cytokines. The levels of these cytokines may be readily assessed using standard assays. For a review of the families of cytokines, see Mosmann & Coffman, *Ann. Rev. Immunol.* 7:145-173 (1989).

Preferred adjuvants for use in eliciting a predominantly Th1-type response
5 include, for example, a combination of monophosphoryl lipid A, preferably 3-de-O-acylated monophosphoryl lipid A (3D-MPL), together with an aluminum salt. MPL adjuvants are available from Corixa Corporation (Seattle, WA; *see* US Patent Nos. 4,436,727; 4,877,611; 4,866,034 and 4,912,094). CpG-containing oligonucleotides (in which the CpG dinucleotide is unmethylated) also induce a predominantly Th1 response. Such oligonucleotides are well
10 known and are described, for example, in WO 96/02555, WO 99/33488 and U.S. Patent Nos. 6,008,200 and 5,856,462. Immunostimulatory DNA sequences are also described, for example, by Sato *et al.*, *Science* 273:352 (1996). Another preferred adjuvant comprises a saponin, such as Quil A, or derivatives thereof, including QS21 and QS7 (Aquila Biopharmaceuticals Inc., Framingham, MA); Escin; Digitonin; or *Gypsophila* or
15 *Chenopodium quinoa* saponins. Other preferred formulations include more than one saponin in the adjuvant combinations of the present invention, for example combinations of at least two of the following group comprising QS21, QS7, Quil A, β-escin, or digitonin.

Alternatively the saponin formulations may be combined with vaccine vehicles composed of chitosan or other polycationic polymers, polylactide and polylactide-co-glycolide particles, poly-N-acetyl glucosamine-based polymer matrix, particles composed of polysaccharides or chemically modified polysaccharides, liposomes and lipid-based particles, particles composed of glycerol monoesters, etc. The saponins may also be formulated in the presence of cholesterol to form particulate structures such as liposomes or ISCOMs.
20 Furthermore, the saponins may be formulated together with a polyoxyethylene ether or ester, in either a non-particulate solution or suspension, or in a particulate structure such as a paucilamellar liposome or ISCOM. The saponins may also be formulated with excipients such as Carbopol^R to increase viscosity, or may be formulated in a dry powder form with a powder excipient such as lactose.
25

In one preferred embodiment, the adjuvant system includes the combination of
30 a monophosphoryl lipid A and a saponin derivative, such as the combination of QS21 and 3D-MPL[®] adjuvant, as described in WO 94/00153, or a less reactogenic composition where the QS21 is quenched with cholesterol, as described in WO 96/33739. Other preferred formulations comprise an oil-in-water emulsion and tocopherol. Another particularly

preferred adjuvant formulation employing QS21, 3D-MPL® adjuvant and tocopherol in an oil-in-water emulsion is described in WO 95/17210.

Another enhanced adjuvant system involves the combination of a CpG-containing oligonucleotide and a saponin derivative particularly the combination of CpG and 5 QS21 as disclosed in WO 00/09159. Preferably the formulation additionally comprises an oil in water emulsion and tocopherol.

Other preferred adjuvants include Montanide ISA 720 (Seppic, France), SAF (Chiron, California, United States), ISCOMS (CSL), MF-59 (Chiron), the SBAS series of adjuvants (*e.g.*, SBAS-2, AS2', AS2," SBAS-4, or SBAS6, available from SmithKline 10 Beecham, Rixensart, Belgium), Detox (Corixa, Hamilton, MT), RC-529 (Corixa, Hamilton, MT) and other aminoalkyl glucosaminide 4-phosphates (AGPs), such as those described in pending U.S. Patent Application Serial Nos. 08/853,826 and 09/074,720, the disclosures of which are incorporated herein by reference in their entireties, and polyoxyethylene ether adjuvants such as those described in WO 99/52549A1.

15 Other preferred adjuvants include adjuvant molecules of the general formula (I): HO(CH₂CH₂O)_n-A-R,
wherein, n is 1-50, A is a bond or -C(O)-, R is C₁₋₅₀ alkyl or Phenyl C₁₋₅₀ alkyl.

One embodiment of the present invention consists of a vaccine formulation comprising a polyoxyethylene ether of general formula (I), wherein n is between 1 and 50, 20 preferably 4-24, most preferably 9; the R component is C₁₋₅₀, preferably C₄-C₂₀ alkyl and most preferably C₁₂ alkyl, and A is a bond. The concentration of the polyoxyethylene ethers should be in the range 0.1-20%, preferably from 0.1-10%, and most preferably in the range 0.1-1%. Preferred polyoxyethylene ethers are selected from the following group: 25 polyoxyethylene-9-lauryl ether, polyoxyethylene-9-stearyl ether, polyoxyethylene-8-stearyl ether, polyoxyethylene-4-lauryl ether, polyoxyethylene-35-lauryl ether, and polyoxyethylene-23-lauryl ether. Polyoxyethylene ethers such as polyoxyethylene lauryl ether are described in the Merck index (12th edition: entry 7717). These adjuvant molecules are described in WO 99/52549.

The polyoxyethylene ether according to the general formula (I) above may, if 30 desired, be combined with another adjuvant. For example, a preferred adjuvant combination is preferably with CpG as described in the pending UK patent application GB 9820956.2.

Any vaccine provided herein may be prepared using well known methods that result in a combination of antigen, immune response enhancer and a suitable carrier or excipient. The compositions described herein may be administered as part of a sustained

release formulation (*i.e.*, a formulation such as a capsule, sponge or gel (composed of polysaccharides, for example) that effects a slow release of compound following administration). Such formulations may generally be prepared using well known technology (*see, e.g.*, Coombes *et al.*, *Vaccine* 14:1429-1438 (1996)) and administered by, for example, 5 oral, rectal or subcutaneous implantation, or by implantation at the desired target site. Sustained-release formulations may contain a polypeptide, polynucleotide or antibody dispersed in a carrier matrix and/or contained within a reservoir surrounded by a rate controlling membrane.

Carriers for use within such formulations are biocompatible, and may also be 10 biodegradable; preferably the formulation provides a relatively constant level of active component release. Such carriers include microparticles of poly(lactide-co-glycolide), polyacrylate, latex, starch, cellulose, dextran and the like. Other delayed-release carriers include supramolecular biovectors, which comprise a non-liquid hydrophilic core (*e.g.*, a cross-linked polysaccharide or oligosaccharide) and, optionally, an external layer comprising 15 an amphiphilic compound, such as a phospholipid (*see, e.g.*, U.S. Patent No. 5,151,254 and PCT applications WO 94/20078, WO/94/23701 and WO 96/06638). The amount of active compound contained within a sustained release formulation depends upon the site of implantation, the rate and expected duration of release and the nature of the condition to be treated or prevented.

Any of a variety of delivery vehicles may be employed within pharmaceutical 20 compositions and vaccines to facilitate production of an antigen-specific immune response that targets tumor cells. Delivery vehicles include antigen presenting cells (APCs), such as dendritic cells, macrophages, B cells, monocytes and other cells that may be engineered to be efficient APCs. Such cells may, but need not, be genetically modified to increase the 25 capacity for presenting the antigen, to improve activation and/or maintenance of the T cell response, to have anti-tumor effects *per se* and/or to be immunologically compatible with the receiver (*i.e.*, matched HLA haplotype). APCs may generally be isolated from any of a variety of biological fluids and organs, including tumor and peritumoral tissues, and may be autologous, allogeneic, syngeneic or xenogeneic cells.

Certain preferred embodiments of the present invention use dendritic cells or 30 progenitors thereof as antigen-presenting cells. Dendritic cells are highly potent APCs (Banchereau & Steinman, *Nature* 392:245-251 (1998)) and have been shown to be effective as a physiological adjuvant for eliciting prophylactic or therapeutic antitumor immunity (*see* Timmerman & Levy, *Ann. Rev. Med.* 50:507-529 (1999)). In general, dendritic cells may be

identified based on their typical shape (stellate *in situ*, with marked cytoplasmic processes (dendrites) visible *in vitro*), their ability to take up, process and present antigens with high efficiency and their ability to activate naïve T cell responses. Dendritic cells may, of course, be engineered to express specific cell-surface receptors or ligands that are not commonly 5 found on dendritic cells *in vivo* or *ex vivo*, and such modified dendritic cells are contemplated by the present invention. As an alternative to dendritic cells, secreted vesicles antigen-loaded dendritic cells (called exosomes) may be used within a vaccine (*see Zitvogel et al., Nature Med.* 4:594-600 (1998)).

Dendritic cells and progenitors may be obtained from peripheral blood, bone 10 marrow, tumor-infiltrating cells, peritumoral tissues-infiltrating cells, lymph nodes, spleen, skin, umbilical cord blood or any other suitable tissue or fluid. For example, dendritic cells may be differentiated *ex vivo* by adding a combination of cytokines such as GM-CSF, IL-4, IL-13 and/or TNF α to cultures of monocytes harvested from peripheral blood. Alternatively, CD34 positive cells harvested from peripheral blood, umbilical cord blood or bone marrow 15 may be differentiated into dendritic cells by adding to the culture medium combinations of GM-CSF, IL-3, TNF α , CD40 ligand, LPS, flt3 ligand and/or other compound(s) that induce differentiation, maturation and proliferation of dendritic cells.

Dendritic cells are conveniently categorized as “immature” and “mature” cells, which allows a simple way to discriminate between two well characterized phenotypes. 20 However, this nomenclature should not be construed to exclude all possible intermediate stages of differentiation. Immature dendritic cells are characterized as APC with a high capacity for antigen uptake and processing, which correlates with the high expression of Fc γ receptor and mannose receptor. The mature phenotype is typically characterized by a lower expression of these markers, but a high expression of cell surface molecules responsible for T 25 cell activation such as class I and class II MHC, adhesion molecules (*e.g.*, CD54 and CD11) and costimulatory molecules (*e.g.*, CD40, CD80, CD86 and 4-1BB).

APCs may generally be transfected with a polynucleotide encoding a protein (or portion or other variant thereof) such that the polypeptide, or an immunogenic portion thereof, is expressed on the cell surface. Such transfection may take place *ex vivo*, and a 30 composition or vaccine comprising such transfected cells may then be used for therapeutic purposes, as described herein. Alternatively, a gene delivery vehicle that targets a dendritic or other antigen presenting cell may be administered to a patient, resulting in transfection that occurs *in vivo*. *In vivo* and *ex vivo* transfection of dendritic cells, for example, may generally

D E E E E E / P A T E M

be performed using any methods known in the art, such as those described in WO 97/24447, or the gene gun approach described by Mahvi *et al.*, *Immunology and Cell Biology* 75:456-460 (1997). Antigen loading of dendritic cells may be achieved by incubating dendritic cells or progenitor cells with the polypeptide, DNA (naked or within a plasmid vector) or RNA; or
5 with antigen-expressing recombinant bacterium or viruses (*e.g.*, vaccinia, fowlpox, adenovirus or lentivirus vectors). Prior to loading, the polypeptide may be covalently conjugated to an immunological partner that provides T cell help (*e.g.*, a carrier molecule). Alternatively, a dendritic cell may be pulsed with a non-conjugated immunological partner, separately or in the presence of the polypeptide.

10 Vaccines and pharmaceutical compositions may be presented in unit-dose or multi-dose containers, such as sealed ampoules or vials. Such containers are preferably hermetically sealed to preserve sterility of the formulation until use. In general, formulations may be stored as suspensions, solutions or emulsions in oily or aqueous vehicles. Alternatively, a vaccine or pharmaceutical composition may be stored in a freeze-dried
15 condition requiring only the addition of a sterile liquid carrier immediately prior to use.

XI. DIAGNOSTIC KITS

The present invention further provides kits for use within any of the above diagnostic methods. Such kits typically comprise two or more components necessary for performing a diagnostic assay. Components may be compounds, reagents, containers and/or equipment. For example, one container within a kit may contain a monoclonal antibody or fragment thereof that specifically binds to a protein. Such antibodies or fragments may be provided attached to a support material, as described above. One or more additional containers may enclose elements, such as reagents or buffers, to be used in the assay. Such kits may also, or alternatively, contain a detection reagent as described above that contains a reporter group suitable for direct or indirect detection of antibody binding.
25

Alternatively, a kit may be designed to detect the level of mRNA encoding a protein in a biological sample. Such kits generally comprise at least one oligonucleotide probe or primer, as described above, that hybridizes to a polynucleotide encoding a protein. Such an oligonucleotide may be used, for example, within a PCR or hybridization assay.
30 Additional components that may be present within such kits include a second oligonucleotide and/or a diagnostic reagent or container to facilitate the detection of a polynucleotide encoding a protein of the invention.

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to one of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

XII. EXAMPLES

The following examples are provided by way of illustration only and not by way of limitation. Those of skill in the art will readily recognize a variety of noncritical parameters that could be changed or modified to yield essentially similar results.

Example 1: Recombinant Fusion Proteins of *M. tuberculosis* Antigens Exhibit Increased Serological Sensitivity

15 A. Materials and Methods

1. Construction of vectors encoding fusion proteins: TbF14

TbF14 is a fusion protein of the amino acid sequence encoding the MTb81 antigen fused to the amino acid sequence encoding the Mo2 antigen. A sequence encoding Mo2 was PCR amplified with the following primers: PDM-294 (T_m 64°C) CGTAATCACGTGCAGAAGTACGGCGGATC (SEQ ID NO:14) and PDM-295 (T_m 63°C) CCGACTAGAATTCACTATTGACAGGCCCATC (SEQ ID NO:15).

DNA amplification was performed using 10 μ l 10X Pfu buffer, 1 μ l 10 mM dNTPs, 2 μ l each of the PCR primers at 10 μ M concentration, 83 μ l water, 1.5 μ l Pfu DNA polymerase (Stratagene, La Jolla, CA) and 50 ng DNA template. For Mo2 antigen, denaturation at 96°C was performed for 2 min; followed by 40 cycles of 96°C for 20 sec, 63°C for 15 sec and 72°C for 2.5 min; and finally by 72°C for 5 min.

A sequence encoding MTb81 was PCR amplified with the following primers: PDM-268 (T_m 66°C) CTAAGTAGTACTGATCGCGTGTGGTGGGC (SEQ ID NO:16) and PDM-296 (T_m 64°C) CATCGATAGGCCTGGCCGCATCGTCACC (SEQ ID NO:17).

30 The amplification reaction was performed using the same mix as above, as follows: denaturation at 96°C for 2 min; followed by 40 cycles of 96°C for 20 sec, 65°C for 15 sec, 72°C for 5 min; and finally by 72°C for 5 min.

The Mo2 PCR product was digested with Eco72I (Stratagene, La Jolla CA) and EcoRI (NEB, Beverly, MA). The MTb81 PCR product was digested with FseI and StuI (NEB, Beverly, MA). These two products were then cloned into an expression plasmid (a modified pET28 vector) with a hexahistidine in frame, in a three way ligation that was
5 digested with FseI and EcoRI. The sequences was confirmed, then the expression plasmid was transformed into the BL21pLysE *E. coli* strain (Novagen, Madison, WI) for expression of the recombinant protein.

2. Construction of vectors encoding fusion proteins: TbF15

TbF15 is a fusion of antigens Ra3, 38 kD (with an N-terminal cysteine), 38-1, 10 and FL TbH4 from *Mycobacterium tuberculosis*, as was prepared a follows. TbF15 was made using the fusion constructs TbF6 and TbF10.

TbF6 was made as follows (see PCT/US99/03268 and PCT/US99/03265).

First, the FL (full-length) TbH4 coding region was PCR amplified with the following primers: PDM-157 CTAGTTAGTACTCAGTCGCAGACCGTG (SEQ ID NO:18) (T_m 61°C) 15 and PDM-160 GCAGTGACGAATTCACCTCGACTCC (SEQ ID NO:19) (T_m 59°C), using the following conditions: 10 μ l 10X Pfu buffer, 1 μ l 10 mM dNTPs, 2 μ l 10 μ M each oligo, 82 μ l sterile water, 1.5 μ l Accuzyme (ISC, Kaysville, UT), 200 ng *Mycobacterium tuberculosis* genomic DNA. Denaturation at 96°C was performed for 2 minutes; followed by 40 cycles of 96°C for 20 seconds, 61°C 15 seconds, and 72°C 5 minutes; and finally by 72°C 20 10 minutes.

The PCR product was digested with ScaI and EcoRI and cloned into pET28Ra3/38kD/38-1A, described below, which was digested with DraI and EcoRI.

pET28Ra3/38kD/38-1A was made by inserting a DraI site at the end of 38-1 before the stop codon using the following conditions. The 38-1 coding region was PCR 25 amplified with the following primers: PDM-69 GGATCCAGCGCTGAGATGAAGACCGATGCCGCT (SEQ ID NO:19) (T_m 68°C) and PDM-83 GGATATCTGCAGAACATTCAAGTTAAAGCCCATTGCGA (SEQ ID NO:20) (T_m 64°C), using the following conditions: 10 μ l 10X Pfu buffer, 1 μ l 10 mM dNTPs, 2 μ l 10 μ M each oligo, 82 μ l sterile water, 1.5 μ l Accuzyme (ISC, Kaysville, UT), 50 ng plasmid 30 DNA. Denaturation at 96°C was performed for 2 minutes; followed by forty cycles of 96°C for 20 seconds, 66°C for 15 seconds and 72°C for 1 minute 10 seconds; and finally 72°C 4 minutes.

The 38-1 PCR product was digested with Eco47III and EcoRI and cloned into the pT7ΔL2Ra3/38kD construct (described in WO/9816646 and WO/9816645) which was digested with EcoRI and Eco47III. The correct construct was confirmed through sequence analysis. The Ra3/38kD/38-1A coding region was then subcloned into pET28 His (a modified pET28 vector) at the NdeI and EcoRI sites. The correct construct (called TbF6) was confirmed through sequence analysis.

Fusion construct TbF10, which replaces the N-terminal cysteine of 38 kD, was made as follows. To replace the cysteine residue at the N-terminus, the 38kD-38-1 coding region from the TbF fusion (described in WO/9816646 and WO/9816645) was amplified using the following primers: PDM-192 TGTGGCTCGAAACCACCGAGCGGTT (SEQ ID NO:21) (T_m 64°C) and PDM-60 GAGAGAATTCTCAGAAGCCCATTGCGAGGACA (SEQ ID NO:22) (T_m 64°C), using the following conditions: 10 μ l 10X Pfu buffer, 1 μ l 10 mM dNTPs, 2 μ l 10 μ M each oligo, 83 μ l sterile water, 1.5 μ l Pfu DNA polymerase (Stratagene, La Jolla, CA), and 50 ng plasmid TbF DNA. The amplification reaction was performed as follows: 96°C for 2 minutes; followed by 40 cycles of 96°C for 20 seconds, 64°C 15 seconds, and 72°C 4 minutes; and finally 72°C 4 minutes. Digest the PCR product with Eco RI and clone into pT7ΔL2Ra3 which has been digested with Stu I and Eco RI. Digest the resulting construct with Nde I and EcoRI and clone into pET28 at those sites. The resulting clone (called TbF10) will be TBF + a cysteine at the 5' end of the 38kD coding region. Transform into BL21 and HMS 174 with pLys S.

The pET28TbF6 (TbF6, described above) construct was digested with StuI (NEB, Beverly, MA) and EcoRI, which released a 1.76 kb insert containing the very back portion of the 38 kD/38-1/FL TbH4 fusion region. This insert was gel purified. The pET28TbF10 construct (TbF10, described above) was digested with the same enzymes and the vector backbone, consisting of 6.45 kb containing the his-tag, the Ra3 coding region and most of the Δ38kD coding region. This insert was gel purified. The insert and vector were ligated and transformed. The correct construct, called TbF15, was confirmed through sequence analysis, then transformed into the BL21 pLysS *E. coli* strain (Novagen, Madison WI). This fusion protein contained the original Cys at the amino terminus of the 38 kD protein.

B. Expression of fusion proteins

1. Expression of fusion proteins

The recombinant proteins were expressed in *E. coli* with six histidine residues at the amino-terminal portion using the pET plasmid vector and a T7 RNA polymerase expression system (Novagen, Madison, WI). *E. coli* strain BL21 (DE3) pLysE (Novagen) was used for high level expression. The recombinant (His-Tag) fusion proteins were purified from the soluble supernatant or the insoluble inclusion body of 1 L of IPTG induced batch cultures by affinity chromatography using the one step QIAexpress Ni-NTA Agarose matrix (QIAGEN, Chatsworth, CA) in the presence of 8M urea.

10 Briefly, 20 ml of an overnight saturated culture of BL21 containing the pET construct was added into 1 L of 2x YT media containing 30 µg/ml kanamycin and 34 µg/ml chloramphenicol, grown at 37°C with shaking. The bacterial cultures were induced with 1 mM IPTG at an OD 560 of 0.3 and grown for an additional 3 h (OD = 1.3 to 1.9). Cells were harvested from 1 L batch cultures by centrifugation and resuspended in 20 ml of binding
15 buffer (0.1 M sodium phosphate, pH 8.0; 10 mM Tris-HCl, pH 8.0) containing 2 mM PMSF and 20 µg/ml leupeptin plus one complete protease inhibitor tablet (Boehringer Mannheim) per 25 ml. *E. coli* was lysed by freeze-thaw followed by brief sonication, then spun at 12 k rpm for 30 min to pellet the inclusion bodies.

20 The inclusion bodies were washed three times in 1% CHAPS in 10 mM Tris-HCl (pH 8.0). This step greatly reduced the level of contaminating LPS. The inclusion body was finally solubilized in 20 ml of binding buffer containing 8 M urea or 8M urea was added directly into the soluble supernatant. Recombinant fusion proteins with His-Tag residues were batch bound to Ni-NTA agarose resin (5 ml resin per 1 L inductions) by rocking at room temperature for 1 h and the complex passed over a column. The flow through was
25 passed twice over the same column and the column washed three times with 30 ml each of wash buffer (0.1 M sodium phosphate and 10 mM Tris-HCl, pH 6.3) also containing 8 M urea. Bound protein was eluted with 30 ml of 150 mM imidazole in wash buffer and 5 ml fractions collected. Fractions containing each recombinant fusion protein were pooled, dialyzed against 10 mM Tris-HCl (pH 8.0) bound one more time to the Ni-NTA matrix,
30 eluted and dialyzed in 10 mM Tris-HCl (pH 7.8). The yield of recombinant protein varies from 25-150 mg per liter of induced bacterial culture with greater than 98% purity. Recombinant proteins were assayed for endotoxin contamination using the *Limulus* assay (BioWhittaker) and were shown to contain < 100 E.U./mg.

DRAFT DUE 2/28/2016

2. Serological assays

ELISA assays were performed with TbF15 using methods known to those of skill in the art, with 200 ng/well of antigen. ELISA assays are performed with TbF14 using methods known to those of skill in the art, with 200 ng/well of antigen.

5 3. Results

The TbF15 fusion protein containing TbRa3, 38kD (with N terminal cysteine), Tb38-1, and full length (FL) TbH4 as described above was used as the solid phase antigen in ELISA. The ELISA protocol is as described above. The fusion recombinant was coated at 200 ng/well. A panel of sera were chosen from a group of TB patients that had previously 10 been shown by ELISA to be positive or borderline positive with these antigens. Such a panel enabled the direct comparison of the fusions with and without the cysteine residue in the 38 kD component. The data are outlined in Figure 5. A total of 23 TB sera were studied and of these 20/23 were detected by TbF6 versus 22/23 for TbF15. Improvements in reactivity were seen in the low reactive samples when TbF15 was used.

15 One of skill in the art will appreciate that the order of the individual antigens within each fusion protein may be changed and that comparable activity would be expected provided that each of the epitopes is still functionally available. In addition, truncated forms of the proteins containing active epitopes may be used in the construction of fusion proteins.

20 Example 2: Cloning, construction, and expression of HTCC#1 full-length, overlapping halves, and deletions as fusion constructs

HTCC#1 (aka MTb40) was cloned by direct T cell expression screening using a T cell line derived from a healthy PPD positive donor to directly screen an *E. coli* based MTb expression library.

A. **Construction and screening of the plasmid expression library**

25 Genomic DNA from *M. tuberculosis* Erdman strain was randomly sheared to an average size of 2 kb and blunt ended with Klenow polymerase, before EcoRI adaptors were added. The insert was subsequently ligated into the 1 screen phage vector and packaged *in vitro* using the PhageMaker extract (Novagen). The phage library (Erd 1 screen) was amplified and a portion was converted into a plasmid expression library. Conversion from 30 phage to plasmid (phagemid) library was performed as follows: the Erd 1 Screen phage library was converted into a plasmid library by autosubcloning using the *E. coli* host strain BM25.8 as suggested by the manufacturer (Novagen). Plasmid DNA was purified from BM25.8 cultures containing the pSCREEN recombinants and used to transform competent

cells of the expressing host strain BL21(DE3)pLysS. Transformed cells were aliquoted into 96 well micro titer plates with each well containing a pool size of ~50 colonies. Replica plates of the 96 well plasmid library format were induced with IPTG to allow recombinant protein expression. Following induction, the plates were centrifuged to pellet the *E. coli* and the bacterial pellet was resuspended in 200 μ l of 1X PBS.

Autologous dendritic cells were subsequently fed with the *E. coli*, washed and exposed to specific T cell lines in the presence of antibiotics to inhibit the bacterial growth. T cell recognition was detected by proliferation and/or production of IFN- γ . Wells that score positive were then broken down using the same protocol until a single clone could be detected. The gene was then sequenced, sub-cloned, expressed and the recombinant protein evaluated.

B. Expression in *E. coli* of the full-length and overlapping constructs of HTCC#1

One of the identified positive wells was further broken down until a single
15 reactive clone (HTCC#1) was identified. Sequencing of the DNA insert followed by search
of the Genebank database revealed a 100% identity to sequences within the *M. tuberculosis*
locus MTCY7H7B (gene identification MTCY07H7B.06) located on region B of the cosmid
clone SCY07H7. The entire open reading frame is ~1,200 bp long and codes for a 40 kDa
(392 amino acids) protein (Fig. 1; HTCC#1 FL). Oligonucleotide PCR primers [5'(5'-CAA
20 TTA CAT ATG CAT CAC CAT CAC CAT CAC ATG AGC AGA GCG TTC ATC ATC-3')
and 3' (5'-CAT GGA ATT CGC CGT TAG ACG ACG TTT CGT A-3')] were designed to
amplify the full-length sequence of HTCC#1 from genomic DNA of the virulent Erdman
strain.

The 5' oligonucleotide contained an *Nde I* restriction site preceding the ATG initiation codon (underlined) followed by nucleotide sequences encoding six histidines (**bold**) and sequences derived from the gene (*italic*). The resultant PCR products was digested with *NdeI* and *EcoRI* and subcloned into the pET17b vector similarly digested with *NdeI* and *EcoRI*. Ligation products were initially transformed into *E. coli* XL1-Blue competent cells (Stratagene, La Jolla, CA) and were subsequently transformed into *E. coli* BL-21 (pLys*iE*) host cells (Novagen, Madison, WI) for expression.

C. Expression of the full length and overlapping constructs of HTCC#1

Several attempts to express the full-length sequence of HTCC#1 in *E. coli* failed either because no transformants could be obtained or because the *E. coli* host cells were

lysed following IPTG induction. HTCC#1 is 392 amino acids long and has 3 trans-membrane (TM) domains which are presumably responsible for the lysing of the *E. coli* culture following IPTG induction.

Thus expression of HTCC#1 was initially attempted by constructing two overlapping fragments coding for the amino (residues 1-223; Fig. 2a) and carboxy (residues 184-392; Fig. 2b) halves.

The N-terminal (residues 1-223) fragment containing the first of the 3 putative transmembrane domains killed (lysed) the host cells, while the C-terminal (residues 184-392) half expressed at high levels in the same host cell. Thus the two trans-membrane domains located in the C-terminal half do not appear to be toxic.

The N-terminal fragment, comprising amino acid residues 1-128 (devoid of the transmembrane domain), was therefore engineered for expression in the same pET17b vector system (Fig. 2c). This construct expressed quite well and there was no toxicity associated with the expressing *E. coli* host.

D. Expression in *E. coli* of the full-length HTCC#1 as an TbRa12 fusion construct

Because of problems associated with the expression of full length HTCC#1, we evaluated the utility of an TbRa12 fusion construct for the generation of a fusion protein that would allow for the stable expression of recombinant HTCC#1.

pET17b vector (Novagen) was modified to include TbRa12, a 14 kDa C-terminal fragment of the serine protease antigen MTB32A of *Mycobacterium tuberculosis* (Skeiky *et al.*). For use as an expression vector, the 3' stop codon of the TbRa12 was substituted with an in frame EcoRI site and the N-terminal end was engineered so as to code for six His-tag residues immediately following the initiator Met. This would facilitate a simple one step purification protocol of TbRa12 recombinant proteins by affinity chromatography over Ni-NTA matrix.

Specifically, the C-terminal fragment of antigen MTB32A was amplified by standard PCR methods using the oligonucleotide primers 5'(CAA TTA CAT ATG CAT CAC CAT CAC CAT CAC ACG GCC GCG TCC GAT AAC TTC and 3' (5'-CTA ATC GAA TTC GGC CGG GGG TCC CTC GGC CAA). The 450 bp product was digested with NdeI and EcoRI and cloned into the pET17b expression vector similarly digested with the same enzymes.

Recombinant HTCC#1 was engineered for expression as a fusion protein with TbRa12 by designing oligonucleotide primers to specifically amplify the full length form.

The 5' oligonucleotide contained a thrombin recognition site. The resulting PCR amplified product was digested with EcoRI and subcloned into the EcoRI site of the pET-TbRa12 vector. Following transformation into the *E. coli* host strain (XL1-blue; Stratagene), clones containing the correct size insert were submitted for sequencing in order to identify those that are in frame with the TbRa12 fusion. Subsequently, the DNA of interest (Fig. 3) was transformed into the BL-21 (pLysE) bacterial host and the fusion protein was expressed following induction of the culture with IPTG.

E. Expression in *E. coli* of HTCC#1 with deletions of the trans-membrane domain(s)

Given the prediction that the 3 predicted trans-membrane (TM) domains are responsible for lysing the *E. coli* host following IPTG induction, recombinant constructs lacking the TM domains were engineered for expression in *E. coli*.

1. Recombinant HTCC#1 with deletion of the first TM (Δ TM-1). A deletion construct lacking the first trans-membrane domain (amino acid residues 150-160) was engineered for expression *E. coli* (Fig. 4a). This construct expressed reasonably well and enough (low mg quantities) was purified for *in vitro* studies. This recombinant antigen was comparable in *in vitro* assays to that of the full-length Ra-12-fusion construct.

T-cell epitope mapping of HTCC#1. Because of the generally low level of expression using the Δ TM-1 construct, the design of the final form of HTCC#1 for expression in *E. coli* was based on epitope mapping. The T-cell epitope was mapped using 30 overlapping peptides (Fig. 4b) on PBMC read out (on four PPD+ donors). The data revealed that peptides 8 through 16 (amino acid residues 92-215) were not immunogenic (Fig. 4c).

2. Recombinant HTCC#1 with deletion of all of the TM domains (Δ TM-2):

A deletion construct of HTCC#1 lacking residues 101 to 203 with a predicted molecular weight of 30.4 kDa was engineered for expression in *E. coli*. The full length HTCC#1 is 40 kDa. There was no toxicity associated with this new deletion construct and the expression level was higher than that of the Δ TM-1 construct (Fig. 4d).

F. Fusion constructs of HTCC#1 and TbH9:

Fig. 5 shows a sequence of HTCC#1 (184-392)-TbH9-HTCC#1 (1-129)

Fig. 6 shows a sequence of HTCC#1 (1-149)-TbH9-HTCC#1 (161-392)

Fig. 7 shows a sequence of HTCC#1 (184-392)-TbH9-HTCC#1 (1-200)

One of skill in the art will appreciate that the order of the individual antigens within each fusion protein may be changed and that comparable activity would be expected provided that each of the epitopes is still functionally available. In addition, truncated forms of the proteins containing active epitopes may be used in the construction of fusion proteins.

5

From the foregoing, it will be appreciated that, although specific embodiments of the invention have been described herein for the purpose of illustration, various modifications may be made without deviating from the spirit and scope of the invention.

RECEIVED
U.S. PATENT AND TRADEMARK OFFICE

WHAT IS CLAIMED IS:

1 1. A pharmaceutical composition comprising an MTb81 antigen or an
2 immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex,
3 and an Mo2 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of
4 the tuberculosis complex.

1 2. The composition of claim 1, wherein the antigens are covalently
2 linked, thereby forming a fusion polypeptide.

1 3. The composition of claim 2, wherein the fusion polypeptide has the
2 amino acid sequence of TbF14.

1 4. A pharmaceutical composition comprising a TbRa3 antigen or an
2 immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, a
3 38kD antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the
4 tuberculosis complex, a Tb38-1 antigen or an immunogenic fragment thereof from a
5 *Mycobacterium* species of the tuberculosis complex, and a FL TbH4 antigen or an
6 immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

1 5. The composition of claim 4, wherein the antigens are covalently
2 linked, thereby forming a fusion polypeptide.

1 6. The composition of claim 5, wherein the fusion polypeptide has the
2 amino acid sequence of TbF15.

1 7. A pharmaceutical composition comprising an HTCC#1 antigen or an
2 immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex,
3 and a TbH9 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of
4 the tuberculosis complex.

1 8. The composition of claim 7, wherein the antigens are covalently
2 linked, thereby forming a fusion polypeptide.

1 9. The composition of claim 7, comprising a full-length HTCC#1 antigen
2 from a *Mycobacterium* species of the tuberculosis complex, and a full-length TbH9 antigen
3 from a *Mycobacterium* species of the tuberculosis complex.

1 10. The composition of claim 9, wherein the antigens are covalently
2 linked, thereby forming a fusion polypeptide.

1 11. The composition of claim 10, wherein the fusion polypeptide has the
2 amino acid sequence of HTCC#1(FL)-TbH9(FL).

1 12. The composition of claim 7, comprising a polypeptide comprising
2 amino acids 184-392 of an HTCC#1 antigen from a *Mycobacterium* species of the
3 tuberculosis complex, a TbH9 antigen or an immunogenic fragment thereof from a
4 *Mycobacterium* species of the tuberculosis complex, and a polypeptide comprising amino
5 acids 1-129 of an HTCC#1 antigen from a *Mycobacterium* species of the tuberculosis
6 complex.

1 13. The composition of claim 12, wherein the antigens are covalently
2 linked, thereby forming a fusion polypeptide.

1 14. The composition of claim 13, wherein the fusion polypeptide has the
2 amino acid sequence of HTCC#1(184-392)/TbH9/HTCC#1(1-129).

1 15. A pharmaceutical composition comprising a TbRa12 antigen or an
2 immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex,
3 and an HTCC#1 antigen or an immunogenic fragment thereof from a *Mycobacterium* species
4 of the tuberculosis complex.

1 16. The composition of claim 15, wherein the antigens are covalently
2 linked, thereby forming a fusion polypeptide.

1 17. The composition of claim 16, wherein the fusion polypeptide has the
2 amino acid sequence of TbRa12-HTCC#1.

1 18. A pharmaceutical composition comprising at least two heterologous
2 antigens from a *Mycobacterium* species of the tuberculosis complex or an immunogenic
3 fragment thereof, wherein the antigen or immungenic fragment thereof is selected from the
4 group consisting of MTb81, Mo2, TbRa3, 38kD, Tb38-1 (MTb11), FL TbH4, HTCC#1
5 (Mtb40), TbH9, MTCC#2 (Mtb41), DPEP, DPPD, TbRa35, TbRa12, MTb59, MTb82, Erd14

6 (Mtb16), FL TbRa35 (Mtb32A), DPV (Mtb8.4), MSL (Mtb9.8), MTI (Mtb9.9A, also known
7 as MTI-A), ESAT-6, α -crystalline, and 85 complex.

1 19. The composition of claim 18, wherein the antigens are covalently
2 linked, thereby forming a fusion polypeptide.

1 20. The composition of claim 1, 4, 7, 15, or 18, wherein the antigens are
2 covalently linked via a chemical linker.

1 21. The composition of claim 20, wherein the chemical linker is an amino
2 acid linker.

1 22. The composition of claim 1, 4, 7, 15, or 18, further comprising at least
2 one additional antigen from a *Mycobacterium* species of the tuberculosis complex, wherein
3 the antigen is selected from the group consisting of MTb81, Mo2, TbRa3, 38kD, Tb38-1
4 (MTb11), FL TbH4, HTCC#1 (Mtb40), TbH9, MTCC#2 (Mtb41), DPEP, DPPD, TbRa35,
5 TbRa12, MTb59, MTb82, Erd14 (Mtb16), FL TbRa35 (Mtb32A), DPV (Mtb8.4), MSL
6 (Mtb9.8), MTI (Mtb9.9A, also known as MTI-A), ESAT-6, α -crystalline, and 85 complex, or
7 an immunogenic fragment thereof.

1 23. The composition of claim 1, 4, 7, 15, or 18, further comprising an
2 adjuvant.

1 24. The composition of claim 23, wherein the adjuvant comprises QS21
2 and MPL.

1 25. The composition of claim 23, wherein the adjuvant is selected from the
2 group consisting of AS2, ENHANZYN, MPL, QS21, CWS, TDM, AGP, CPG, Leif, saponin,
3 and saponin mimetics.

1 26. The composition of claim 1, 4, 7, 15, or 18, further comprising BCG.

1 27. The composition of claim 1, 4, 7, 15, or 18, further comprising an NS1
2 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the
3 tuberculosis complex.

1 28. The composition of claim 1, 4, 7, 15, or 18, wherein the
2 *Mycobacterium* species is *Mycobacterium tuberculosis*.

1 29. An expression cassette comprising a nucleic acid encoding an MTb81
2 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the
3 tuberculosis complex, and a nucleic acid encoding an Mo2 antigen or an immunogenic
4 fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

1 30. The expression cassette of claim 29, wherein the nucleic acid encodes
2 a fusion polypeptide comprising an MTb81 antigen or an immunogenic fragment thereof and
3 a nucleic acid encoding an Mo2 antigen or an immunogenic fragment thereof.

1 31. The expression cassette of claim 30, wherein the nucleic acid encodes
2 a fusion polypeptide having the amino acid sequence of TbF14.

1 32. The expression cassette of claim 31, wherein the nucleic acid has the
2 nucleotide sequence of the nucleic acid encoding TbF14.

1 33. An expression cassette comprising a nucleic acid encoding a TbRa3
2 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the
3 tuberculosis complex, a nucleic acid encoding a 38kD antigen or an immunogenic fragment
4 thereof from a *Mycobacterium* species of the tuberculosis complex, a nucleic acid encoding a
5 Tb38-1 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the
6 tuberculosis complex, and a nucleic acid encoding a FL TbH4 antigen or an immunogenic
7 fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

1 34. The expression cassette of claim 33, wherein the nucleic acid encodes
2 a fusion polypeptide comprising a TbRa3 antigen or an immunogenic fragment thereof, a
3 38kD antigen or an immunogenic fragment thereof, a Tb38-1 antigen or an immunogenic
4 fragment thereof, and a nucleic acid encoding a FL TbH4 antigen or an immunogenic
5 fragment thereof.

1 35. The expression cassette of claim 34, wherein the nucleic acid encodes
2 a fusion polypeptide having the amino acid sequence of TbF15.

1 36. The expression cassette of claim 35, wherein the nucleic acid has the
2 nucleotide sequence of the nucleic acid encoding TbF15.

1 37. An expression cassette comprising a nucleic acid encoding an
2 HTCC#1 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the
3 tuberculosis complex, and a nucleic acid encoding a TbH9 antigen or an immunogenic
4 fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

1 38. The expression cassette of claim 37, comprising a nucleic acid
2 encoding a full-length HTCC#1 antigen from a *Mycobacterium* species of the tuberculosis
3 complex, and a nucleic acid encoding a full-length TbH9 antigen from a *Mycobacterium*
4 species of the tuberculosis complex.

1 39. The expression cassette of claim 37, comprising a nucleic acid
2 encoding a polypeptide comprising amino acids 184-392 of an HTCC#1 antigen from a
3 *Mycobacterium* species of the tuberculosis complex, a nucleic acid encoding a TbH9 antigen
4 or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis
5 complex, and a nucleic acid encoding a polypeptide comprising amino acids 1-129 of an
6 HTCC#1 antigen from a *Mycobacterium* species of the tuberculosis complex.

1 40. The expression cassette of claim 37, wherein the nucleic acid encodes
2 a fusion polypeptide comprising an HTCC#1 antigen or an immunogenic fragment thereof,
3 and a TbH9 antigen or an immunogenic fragment thereof.

1 41. The expression cassette of claim 38, wherein the nucleic acid encodes
2 a fusion polypeptide comprising a full-length HTCC#1 antigen, and a full-length TbH9
3 antigen.

1 42. The expression cassette of claim 39, wherein the nucleic acid encodes
2 a fusion polypeptide comprising amino acids 184-392 of an HTCC#1, a TbH9 antigen or an
3 immunogenic fragment thereof, and amino acids 1-129 of an HTCC#1 antigen.

1 43. The expression cassette of claim 41, wherein the nucleic acid encodes
2 a fusion polypeptide having the amino acid sequence of HTCC#1(FL)-TbH9(FL).

1 44. The expression cassette of claim 43, wherein the nucleic acid has the
2 nucleotide sequence of the nucleic acid encoding HTCC#1(FL)-TbH9(FL).

1 45. The expression cassette of claim 42, wherein the nucleic acid encodes
2 a fusion polypeptide having the amino acid sequence of HTCC#1(184-
3 392)/TbH9/HTCC#1(1-129).

1 46. The expression cassette of claim 45, wherein the nucleic acid has the
2 nucleotide sequence of the nucleic acid encoding HTCC#1(184-392)/TbH9/HTCC#1(1-129).

1 47. An expression cassette comprising a nucleic acid encoding a TbRa12
2 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the
3 tuberculosis complex, and a nucleic acid encoding an HTCC#1 antigen or an immunogenic
4 fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

1 48. The expression cassette of claim 47, wherein the nucleic acid encodes
2 a fusion polypeptide comprising an Ra12 antigen or an immunogenic fragment thereof, and
3 an HTCC#1 antigen or an immunogenic fragment thereof.

1 49. The expression cassette of claim 48, wherein the nucleic acid encodes
2 a fusion polypeptide having the amino acid sequence of TbRa12-HTCC#1.

1 50. The expression cassette of claim 49, wherein the nucleic acid has the
2 nucleotide sequence of the nucleic acid encoding TbRa12-HTCC#1.

1 51. An expression cassette comprising a nucleic acid encoding at least two
2 heterologous antigens from a *Mycobacterium* species of the tuberculosis complex or an
3 immunogenic fragment thereof, wherein the antigen or immunogenic fragment thereof is
4 selected from the group consisting of MTb81, Mo2, TbRa3, 38kD, Tb38-1 (MTb11), FL
5 TbH4, HTCC#1 (Mtb40), TbH9, MTCC#2 (Mtb41), DPEP, DPPD, TbRa35, TbRa12,
6 MTb59, MTb82, Erd14 (Mtb16), FL TbRa35 (Mtb32A), DPV (Mtb8.4), MSL (Mtb9.8), MTI
7 (Mtb9.9A, also known as MTI-A), ESAT-6, α -crystalline, and 85 complex.

1 52. The expression cassette of claim 51, wherein the nucleic acid encodes
2 a fusion polypeptide.

1 53. The expression cassette of claim 29, 33, 37, 47 or 51, further
2 comprising a nucleic acid encoding at least one additional antigen from a *Mycobacterium*
3 species of the tuberculosis complex, wherein the antigen is selected from the group consisting

4 of MTb81, Mo2, TbRa3, 38kD, Tb38-1 (MTb11), FL TbH4, HTCC#1 (Mtb40), TbH9,
5 MTCC#2 (Mtb41), DPEP, DPPD, TbRa35, TbRa12, MTb59, MTb82, Erd14 (Mtb16), FL
6 TbRa35 (Mtb32A), DPV (Mtb8.4), MSL (Mtb9.8), MTI, ESAT-6, α -crystalline, and 85
7 complex, or an immunogenic fragment thereof.

1 54. The expression cassette of claim 29, 33, 37, 47 or 51, further
2 comprising a nucleic acid encoding an NS1 antigen or an antigenic fragment thereof from a
3 *Mycobacterium* species of the tuberculosis complex.

1 55. The expression cassette of claim 29, 33, 37, 47 or 51, wherein the
2 *Mycobacterium* species is *Mycobacterium tuberculosis*.

1 56. A method for eliciting an immune response in a mammal, the method
2 comprising the step of administering to the mammal an immunologically effective amount of
3 a pharmaceutical composition comprising an MTb81 antigen or an immunogenic fragment
4 thereof from a *Mycobacterium* species of the tuberculosis complex, and an Mo2 antigen or an
5 immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

1 57. The method of claim 56, wherein the antigens are covalently linked,
2 thereby forming a fusion polypeptide.

1 58. The method of claim 57, wherein the fusion polypeptide has the amino
2 acid sequence of TbF14.

1 59. A method for eliciting an immune response in a mammal, the method
2 comprising the step of administering to the mammal an immunologically effective amount of
3 a pharmaceutical composition comprising a TbRa3 antigen or an immunogenic fragment
4 thereof from a *Mycobacterium* species of the tuberculosis complex, a 38kD antigen or an
5 immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, a
6 Tb38-1 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the
7 tuberculosis complex, and a FL TbH4 antigen or an immunogenic fragment thereof from a
8 *Mycobacterium* species of the tuberculosis complex.

1 60. The method of claim 59, wherein the antigens are covalently linked,
2 thereby forming a fusion polypeptide.

1 61. The method of claim 60, wherein the fusion polypeptide has the amino
2 acid sequence of TbF15.

1 62. A method for eliciting an immune response in a mammal, the method
2 comprising the step of administering to the mammal an immunologically effective amount of
3 a pharmaceutical composition comprising an HTCC#1 antigen or an immunogenic fragment
4 thereof from a *Mycobacterium* species of the tuberculosis complex, and a TbH9 antigen or an
5 immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

1 63. The method of claim 62, wherein the pharmaceutical composition
2 comprises a full-length HTCC#1 antigen from a *Mycobacterium* species of the tuberculosis
3 complex, and a full-length TbH9 antigen from a *Mycobacterium* species of the tuberculosis
4 complex.

1 64. The method of claim 63, wherein the antigens are covalently linked,
2 thereby forming a fusion polypeptide.

1 65. The method of claim 64, wherein the fusion polypeptide has the amino
2 acid sequence of HTCC#1(FL)-TbH9(FL).

1 66. The method of claim 62, wherein the pharmaceutical composition
2 comprises a polypeptide comprising amino acids 184-392 of an HTCC#1 antigen from a
3 *Mycobacterium* species of the tuberculosis complex, a TbH9 antigen or an immunogenic
4 fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and a
5 polypeptide comprising amino acids 1-129 of an HTCC#1 antigen from a *Mycobacterium*
6 species of the tuberculosis complex.

1 67. The method of claim 66, wherein the antigens are covalently linked,
2 thereby forming a fusion polypeptide.

1 68. The method of claim 67, wherein the fusion polypeptide has the amino
2 acid sequence of HTCC#1(184-392)/TbH9/HTCC#1(1-129).

1 69. A method for eliciting an immune response in a mammal, the method
2 comprising the step of administering to the mammal an immunologically effective amount of
3 a pharmaceutical composition comprising a TbRa12 antigen or an immunogenic fragment

4 thereof from a *Mycobacterium* species of the tuberculosis complex, and an HTCC#1 antigen
5 or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis
6 complex.

1 70. The method of claim 69, wherein the antigens are covalently linked,
2 thereby forming a fusion polypeptide.

1 71. The method of claim 70, wherein the fusion polypeptide has the amino
2 acid sequence of TbRa12-HTCC#1.

1 72. A method for eliciting an immune response in a mammal, the method
2 comprising the step of administering to the mammal an immunologically effective amount of
3 a pharmaceutical composition comprising at least two heterologous antigens from a
4 *Mycobacterium* species of the tuberculosis complex or an immunogenic fragment thereof,
5 wherein the antigen or immungenic fragment thereof is selected from the group consisting of
6 MTb81, Mo2, TbRa3, 38kD, Tb38-1 (MTb11), FL TbH4, HTCC#1 (MtB40), TbH9,
7 MTCC#2 (MtB41), DPEP, DPPD, TbRa35, TbRa12, MTb59, MTb82, Erd14 (MtB16), FL
8 TbRa35 (MtB32A), DPV (MtB8.4), MSL (MtB9.8), MTI (MtB9.9A, also known as MTI-A),
9 ESAT-6, α -crystalline, and 85 complex.

1 73. The method of claim 72, wherein the antigens are covalently linked,
2 thereby forming a fusion protein.

1 74. The method of claim 56, 59, 62, 69, or 72, wherein the mammal has
2 been immunized with BCG.

1 75. The method of claim 56, 59, 62, 69, or 72, wherein the mammal is a
2 human.

1 76. The method of claim 56, 59, 62, 69, or 72, wherein the composition is
2 administered prophylactically.

1 77. The method of claim 56, 59, 62, 69, or 72, wherein the pharmaceutical
2 composition further comprises an adjuvant.

1 78. The method of claim 77, wherein the adjuvant comprises QS21 and
2 MPL.

1 79. The method of claim 77, wherein the adjuvant is selected from the
2 group consisting of AS2, ENHANZYN, MPL, QS21, CWS, TDM, AGP, CPG, Leif, saponin,
3 and saponin mimetics.

1 80. A method for eliciting an immune response in a mammal, the method
2 comprising the step of administering to the mammal an immunologically effective amount of
3 an expression cassette comprising a nucleic acid encoding an MTb81 antigen or an
4 immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex,
5 and a nucleic acid encoding an Mo2 antigen or an immunogenic fragment thereof from a
6 *Mycobacterium* species of the tuberculosis complex.

1 81. The method of claim 80, wherein the nucleic acid encodes a fusion
2 polypeptide comprising an MTb81 antigen or an immunogenic fragment thereof, and an Mo2
3 antigen or an immunogenic fragment thereof.

1 82. The method of claim 81, wherein the nucleic acid encodes a fusion
2 polypeptide having the amino acid sequence of TbF14.

1 83. The method of claim 82, wherein the nucleic acid has the nucleotide
2 sequence of the nucleic acid encoding TbF14.

1 84. A method for eliciting an immune response in a mammal, the method
2 comprising the step of administering to the mammal an immunologically effective amount of
3 an expression cassette comprising a nucleic acid encoding a TbRa3 antigen or an
4 immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, a
5 nucleic acid encoding a 38kD antigen or an immunogenic fragment thereof from a
6 *Mycobacterium* species of the tuberculosis complex, a nucleic acid encoding a Tb38-1
7 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the
8 tuberculosis complex, and a nucleic acid encoding a FL TbH4 antigen or an immunogenic
9 fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

1 85. The method of claim 84, wherein the nucleic acid encodes a fusion
2 polypeptide comprising a TbRa3 antigen or an immunogenic fragment thereof, a 38kD
3 antigen or an immunogenic fragment thereof, a Tb38-1 antigen or an immunogenic fragment
4 thereof, and a FL TbH4 antigen or an immunogenic fragment thereof.

1 86. The method of claim 85, wherein the nucleic acid encodes a fusion
2 polypeptide having the amino acid sequence of TbF15.

1 87. The method of claim 86, wherein the nucleic acid has the nucleotide
2 sequence of the nucleic acid encoding TbF15.

1 88. A method for eliciting an immune response in a mammal, the method
2 comprising the step of administering to the mammal an immunologically effective amount of
3 an expression cassette comprising a nucleic acid encoding an HTCC#1 antigen or an
4 immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex,
5 and a nucleic acid encoding a TbH9 antigen or an immunogenic fragment thereof from a
6 *Mycobacterium* species of the tuberculosis complex.

1 89. The method of claim 88, wherein the nucleic acid encodes a fusion
2 polypeptide comprising an HTCC#1 antigen or an immunogenic fragment thereof, and a
3 TbH9 antigen or an immunogenic fragment thereof.

1 90. The method of claim 89, wherein the nucleic acid encodes a fusion
2 polypeptide comprising a full-length HTCC#1 antigen or an immunogenic fragment thereof,
3 and a full-length TbH9 antigen or an immunogenic fragment thereof.

1 91. The method of claim 90, wherein the nucleic acid encodes a fusion
2 polypeptide having the amino acid sequence of HTCC#1(FL)-TbH9(FL).

1 92. The method of claim 91, wherein the nucleic acid has the nucleotide
2 sequence of the nucleic acid encoding HTCC#1(FL)-TbH9(FL).

1 93. The method of claim 89, wherein the nucleic acid encodes a fusion
2 polypeptide comprising a polypeptide comprising amino acids 184-392 of an HTCC#1
3 antigen, a TbH9 antigen or an immunogenic fragment thereof, and a polypeptide comprising
4 amino acids 1-129 of an HTCC#1 antigen.

1 94. The method of claim 93, wherein the nucleic acid encodes a fusion
2 polypeptide having the amino acid sequence of HTCC#1(184-392)/TbH9/HTCC#1(1-129).

1 95. The method of claim 93, wherein the nucleic acid has the nucleotide
2 sequence of the nucleic acid encoding HTCC#1(184-392)/TbH9/HTCC#1(1-129).

1 96. A method for eliciting an immune response in a mammal, the method
2 comprising the step of administering to the mammal an immunologically effective amount of
3 an expression cassette comprising a nucleic acid encoding a TbRa12 antigen or an
4 immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex,
5 and a nucleic acid encoding an HTCC#1 antigen or an immunogenic fragment thereof from a
6 *Mycobacterium* species of the tuberculosis complex.

1 97. The method of claim 96, wherein the nucleic acid encodes a fusion
2 polypeptide comprising a TbRa12 antigen or an immunogenic fragment thereof, and an
3 HTCC#1 antigen or an immunogenic fragment thereof.

1 98. The method of claim 97, wherein the nucleic acid encodes a fusion
2 polypeptide having the amino acid sequence of TbRa12-HTCC#1.

1 99. The method of claim 98, wherein the nucleic acid has the nucleotide
2 sequence of the nucleic acid encoding TbRa12-HTCC#1.

1 100. A method for eliciting an immune response in a mammal, the method
2 comprising the step of administering to the mammal an immunologically effective amount of
3 an expression cassette comprising a nucleic acid encoding at least two heterologous antigens
4 from a *Mycobacterium* species of the tuberculosis complex or an immunogenic fragment
5 thereof, wherein the antigen or immunogenic fragment thereof is selected from the group
6 consisting of MTb81, Mo2, TbRa3, 38kD, Tb38-1 (MTb11), FL TbH4, HTCC#1 (Mtb40),
7 TbH9, MTCC#2 (Mtb41), DPEP, DPPD, TbRa35, TbRa12, MTb59, MTb82, Erd14 (Mtb16),
8 FL TbRa35 (Mtb32A), DPV (Mtb8.4), MSL (Mtb9.8), MTI (Mtb9.9A, also known as MTI-
9 A), ESAT-6, α -crystalline, and 85 complex.

1 101. The method of claim 100, wherein the nucleic acid encodes a fusion
2 polypeptide.

1 102. The method of claim 80, 84, 88, 96, or 100, wherein the mammal has
2 been immunized with BCG.

1 103. The method of claim 80, 84, 88, 96, or 100, wherein the mammal is a
2 human.

104. The method of claim 80, 84, 88, 96, or 100, wherein the composition is administered prophylactically.

1 105. A fusion protein comprising an MTb81 antigen or an immunogenic
2 fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and an Mo2
3 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the
4 tuberculosis complex.

1 106. The protein of claim 105, wherein the fusion polypeptide has the
2 amino acid sequence of TbF14.

1 107. A fusion protein comprising a TbRa3 antigen or an immunogenic
2 fragment thereof from a *Mycobacterium* species of the tuberculosis complex, a 38kD antigen
3 or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis
4 complex, a Tb38-1 antigen or an immunogenic fragment thereof from a *Mycobacterium*
5 species of the tuberculosis complex, and a FL TbH4 antigen or an immunogenic fragment
6 thereof from a *Mycobacterium* species of the tuberculosis complex.

1 108. The protein of claim 107, wherein the fusion polypeptide has the
2 amino acid sequence of TbF15.

1 109. A fusion protein comprising an HTCC#1 antigen or an immunogenic
2 fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and a TbH9
3 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the
4 tuberculosis complex.

1 110. The protein of claim 109, comprising a full-length HTCC#1 antigen
2 from a *Mycobacterium* species of the tuberculosis complex, and a full-length TbH9 antigen
3 from a *Mycobacterium* species of the tuberculosis complex.

1 111. The protein of claim 110, wherein the fusion polypeptide has the
2 amino acid sequence of HTCC#1(FL)-TbH9(FL).

1 112. The protein of claim 109, comprising a polypeptide comprising amino
2 acids 184-392 of an HTCC#1 antigen from a *Mycobacterium* species of the tuberculosis
3 complex, a TbH9 antigen or an immunogenic fragment thereof from a *Mycobacterium*

4 species of the tuberculosis complex, and a polypeptide comprising amino acids 1-129 of an
5 HTCC#1 antigen from a *Mycobacterium* species of the tuberculosis complex.

1 113. The protein of claim 112, wherein the fusion polypeptide has the
2 amino acid sequence of HTCC#1(184-392)/TbH9/HTCC#1(1-129).

1 114. A fusion protein comprising a TbRa12 antigen or an immunogenic
2 fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and an
3 HTCC#1 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the
4 tuberculosis complex.

1 115. The protein of claim 114, wherein the fusion polypeptide has the
2 amino acid sequence of TbRa12-HTCC#1.

FUSION PROTEINS OF MYCOBACTERIUM TUBERCULOSIS

ABSTRACT OF THE DISCLOSURE

The present invention relates to fusion proteins containing at least two *Mycobacterium* species antigens. In particular, it relates to nucleic acids encoding fusion proteins that include two or more individual *M. tuberculosis* antigens, which increase serological sensitivity of sera from individuals infected with tuberculosis, and methods for their use in the diagnosis, treatment, and prevention of tuberculosis infection.

SF 1145088 v1

10

卷之三

Figure 1: Nucleotide Sequence of TbF14
Sheet 1 of 4

FEATURES	Location/Qualifiers
misc_feature	5072..5095
	/note="His tag coding region"
misc_feature	5096..7315
	/note="MtB81 coding region"
misc_feature	7316..8594
	/note="Mo2 coding region"

TGGCGAATGGGACGCCCTGTAGCGCGCATTAAGCGGGCGGGTGTGGTGGTTACGCAGCGT
GACCGCTACACTGCCAGCGCCCTAGCGCCGCTCCTTCGCTTCTCCCTTCCTCGCCAC
GTTCGCCGGCTTCCTCGCAAGCTCTAAATCGGGGGCTCCCTTAGGGTCCGATTAGTGCTT
ACGGCACCTCGACCCCCAAAAACTTGATTAGGGTGTGGTACCGTAGTGGGCCATGCCCTGATA
GACGGTTTTCGCCCTTGACGTTGGAGTCCACGTTCTTAATAGTGACTTGTCCAAACTGG
AACAAACACTCAACCCTATCTCGGTCTATTCTTGATTATAAGGGATTTGCCGATTGCCCTA
TTGGTTAAAAATGAGCTGATTTAACAAAATTAAACGGAATTAAACAAAATTAAACGTTAC
AATTCAGGTGGCACTTTGGGGAAATGTGCCGGAACCCCTATTGTTATTTCTAAATACA
TTCAAATATGTATCCGCTCATGAATTAAATTCTAGAAAAACTCATCGAGCATCAAATGAAACTGCA
ATTATTATCATATCAGGATTATCAATACCATATTGGAAAAAGCCGTTCTGTAATGAAGGAGAAA
ACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGCTCGCATTCCGACTCGTCAA
CATCAATACAACCTATTAAATTCCCTCGTCAAAATAAGGTATCAAGTGAGAAATCACCATGAG
TGACGACTGAATCCGGTGAGAATGGCAAAAGTTATGCATTCTTCCAGACTTGTCAACAGGCC
AGCCATTACGCTCGTCATCAAATCACTCGCATCAACCAAACCGTTATTCACTCGTATTGCC
GAGCGAGACGAAATACGCGATCGCTGTTAAAGGACAATTACAAACAGGAATCGAACCGGC
GCAGGAACACTGCCAGCGCATCAACAAATTTCACCTGAATCAGGATATTCTCTAAACCTGGA
ATGCTGTTCCCGGGATCGCAGTGGTAGTAACCATGCATCATCAGGAGTACGGATAAAATGCT
TGATGGTCGGAAGAGGCATAAAATTCCGTAGCCAGTTAGTCTGACCCTCATCTGTAACATCAT
TGGCAACGCTACCTTGCCATGTTAGAAACAAACTCTGGCGATCGGCTCCATACAATCGAT
AGATTGTGCGACCTGATTGCCGACATTATCGCGAGCCCATTACCCATATAAATCAGCATCCA
TGGTGGAAATTAAATCGCGGCCAGAGCAAGACGTTCCCGTTGAATATGGCTCATAACACCCCTG
TATTACTGTTATGTAAGCAGACAGTTATTGTTATGACCAAAATCCCTAACGTGAGTTTCG
TTCCACTGAGCGTAGGCCACCACTCAAGAACCTGAGAAAAGATCAAAGGATCTTGTGAGATCCTTCTGCGC
GTAATCTGCTGCTTGCACAAACAAAAACCCAGCTACCGCGGTGGTTGGACTCAAGAG
CTACCAACTCTTCCGAAGGTAACTGGCTCAGCAGAGCGCAGATAACAAACTGTCTTCTA
GTGTAGCCGTAGTTAGGCCACCACTCAAGAACCTGAGCTTACAGCCCTACATACCTCGCTCTGCTA
ATCCTGTTACCACTGGCTGCTGCCAGTGGCGATAAGTCGTGTTACCGGGTTGGACTCAAGACGA
TAGTTACCGGATAAGCGCAGCGTGGCTGAACGGGGGTTCTGCACACAGCCAGCTGGAG
CGAACGACCTACACCGAACTGAGATAACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTCCGAA
GGGAGAAAAGCGGACAGGTATCCGTAAGCGGAGGGTCGGAACAGGAGAGCGCAGGAGGGAGCTT
CCAGGGGAAACGCCGGTATCTTATAGTCCTGCGGTTGCCACCTCTGACTTGAGCGTCGA
TTTTGTGATGCTCGTCAGGGGGCGAGCCTATGGAAAAACGCCAGCAACGCCCTTTACGG
TTCCTGGCCTTGTGCTGGCTTGTGTCACATGTTCTTCTGCGTTATCCCTGATTCTGTT
AACCGTATTACCGCCTTGAGTGAGCTGATACCGCTGCCGAGCGAACGAGCGCAGCGAG
TCAGTGAGCGAGGAAGCGGAAGAGCGCTGATCGGGTATTTCCTCTTACGCATCTGCGGTATT
TCACACCGCATATGGTGCACCTCAAGTACAATCTGCTGATGCCGATAGTTAACCGAGTATA

Figure 1: Nucleotide Sequence of TbF14
Sheet 2 of 4

CACTCCGCTATCGCTACGTACTGGTCATGGCTGCGCCCGACACCCGCCAACACCCGCTGACGC
GCCCTGACGGGTTGTCTGCTCCCAGCATCCGCTTACAGACAAGCTGTGACCCTCAGGGAGCTG
CATGTGTCAAGAGTTTCAACCGTCATACCGAAACCGCGAGGCAGCTGGCTAAAGCTCATCAGC
GTGGTCGTGAAGCGATTACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTCTCCAG
AAGCGTTAATGTCTGGCTCTGATAAAAGCGGGCATGTTAAGGGCGTTTTCTGTTGGTCAC
TGATGCCTCCGTAAAGGGGATTCTGTTCATGGGGTAATGATAACCGATGAAACGAGAGAGGAT
GCTCACGATAACGGTTACTGATGATGAAACATGCCCGTTACTGGAACGTTGTGAGGGTAAACA
GGCGGTATGGATGCCCGGGACCAGAGAAAAACTCACTCAGGGTCAATGCCAGCGCTCGTTAAC
AGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGAGATCCGAAACATAATGGTGCA
GGCGCTGACTTCCGTTCCAGACTTACGAAACACGGAAACCGAAGACCATTATGTTGC
TCAGGTGCGAGACGTTTGCAGCAGCGACTCGCTCACGTTGCTCGGTATCGGTGATTCAATTCTG
CTAACCAAGTAAGGCAACCCCGCCAGCCTAGCCGGTCCTCAACGACAGGAGCACGATCATGCGAC
CCGTGGGGCGCCATGCCGGGATAATGGCTGCTCTGCCGAAACGTTGGTGGCGGGACCAGT
GACGAAGGCTTGAGCGAGGGCGTCAAGATTCCAATACCGCAAGCGACAGGCCGATCATGTCGC
GCTCCAGCGAAAGCGGTCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGCTTACGAGTTG
CATGATAAAGAAGACAGTCATAAGTGCAGCGACGATAGTCATGCCCGCGCCACCGGAAGGAGCT
GACTGGGTGAAGGCTCTCAAGGGATCGGTGAGATCCCGTGCCTAATGAGTGAGCTAACTTAC
ATTAATTGCGTTGCGCTCACTGCCGTTCCAGTCGGAAACCTGTCGTGCCAGCTGCATTAATG
AATCGGCCAACCGCGGGGAGAGCGGTTGCGTATTGGCGCCAGGGTGGTTTCTTCAACCA
GTGAGACGGCAACAGCTGATTGCCCTCACCGCCTGGCCCTGAGAGAGGTTGAGCAAGCGGTCCA
CGCTGGTTGCCCGAGCAGCGAAAATCCTGTTGATGGTGGTTACGGCGGGATATAACATGAGC
TGTCTCGTATCGTATCCACTACCGAGATATCCGACCAACGCGCAGCCGGACTCGTAA
TGGCGCGCATTGCCAGCGCATCTGATCGTGGCAACCAGCATCGCAGTGGAAACGATGCCCT
CATTCAAGCATTGCAATGGTTGTTGAAAACCGGACATGGCACTCCAGTCGCTTCCGTTCCGCTA
TCGGCTGAATTGATTGCGAGTGAGATATTGAGCTGGTGAACCGGACATGCGACCAGCGCCGAGACAG
AACTTAATGGCCCGCTAACAGCGCATTGCTGGTGAACCGGACATGCGACCAGATGCTCCACGCCA
GTCGCGTACCGTCTCATGGAGAAAATAACTGTTGATGGTGTCTGGTCAAGAGACATCAAGAA
ATAACGCCGAACATTAGTGCAGGCAGCTCCACAGCAATGGCATCCTGGTATCCAGCGGATAGT
TAATGATCAGCCCAC TGACCGTGTGCGGAGAAAGATTGTCACCGCCGCTTACAGGCTTCGACGC
CGCTCGTCTACCACATGACACACCACGCTGGCACCCAGTTGATCGCGAGATTAAATGCCG
CGACAATTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTT
TGGCCGCCAGTTGTTGCGCACCGGGTGGGAATGTAATTGAGCTCCGCACTGCCGCTTCCACTT
TTTCCCGCTTTCGAGAAACGTTGCTGGTGGCTGGTCAACACGGGAAACGGTCTGATAAGAGA
CACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTACATTCAACGCCGCTGAATTGACTCT
CTTCCGGCGCTATCATGCCATACCGCGAAAGGTTTGCCTGCGCATTGATGGTGTCCGGGATCTGA
CGCTCTCCCTATGCGACTCCTGCATTAGGAAGCAGCCAGTAGTAGGTTGAGGCCGTTGAGCACC
GCCGCCGAAGGAATGGTGCATGCAAGGAGATGGCGCCAACAGTCCCCGGCACGGGGCTGCC
ACCACACCGCCAACAGCGCTCATGAGGCCGAAGTGGCGAGGCCGATCTCCCCATCGTG
ATGTCGGCGATATAAGGCGCCAGCAACCGCACCTGTCGGCCGGTGATGCCGGCACGATGCGTCCG
GCGTAGAGGATCGAGATCTGATCCCGAAATTAAACGACTCACTATAGGGATTGTGAGCGG
ATAACAAATTCCCTCTAGAAATAATTGTTAACCTTAAGAAGGAGATACATATGCGACATCA
CCACCATCACCACACTGATCGCGTGTGGCAACTTGCACATCGCTGGGTGCTACGACTT
CGTGAACAATGAAGCCCTGCCTGGCACCGATATGACCCGGACAGCTCTGGCGGGCTGACAA

Figure 1: Nucleotide Sequence of TbF14
Sheet 3 of 4

GTCGTCGCCGACCTGACCCCGAGAACCAAGCTCTGTTGAACGCCGCAGCAGCTGCAGGCCAG
ATCGACAAGTGGCACCGCGCTGGGTGATCGAGCCCATCGACATGGATGCCTACCGCCAGTCCTC
ACCGAGATCGGCTACCTGCTCCGAACCTGATGACTCACCATACCACGTCCGGTGTGACGCT
GAGATCACACGACCGCCGGCCCCAGCTGGTGGTGCCGGTGTCAACGCCGGTTGCTCTGAAC
GCGGCCAACGCTCGTGGGCTCCCTACGACGCCCTGTATGGCACCGATGTCATCCCCGAGACC
GACGGCGCCGAAAAAGGCCAACGTACAACAAGGTTCGTGGCGACAAGGTGATCGCGTATGCCGC
AAGTTCTCGACGACAGTGTCCGCTGTCGTCGGTTCTTGGGACGCCACCGGTTCACAGTG
CAGGATGGCCAGCTCGTGGTGCCTGGGATAAGTCCACCGGCTGGCCAACCCGGCCAGTTC
GCCGGCTACACCGGCGCAGCCAGTCCGACATCGGTGCTGCTAATCAATCACGGTTGCACATC
GAGATCTGATCGATCCGGAGTCGCAAGGTGGCACCACCGACCGGGCGGCTCAAGGACGTGATC
CTGGAATCCGCGATCACCAACGATCATGGACTTCGAGGACTCGGTGGCCGCGTGGACGCCGCCAC
AAGGTGCTGGTTATCGGAACCTGGCTCGGCCTGAACAAGGGCGACCTGGCAGCGGTAGACAAG
GACGGCACCGCTTCTGCGGGTGTCAATAGGGACCGGAACTAACCGCACCCGGCGTGGCCAG
TTCACGCTGCCTGGACGCAGCCTCATGTTCGTCCGCAACGTCGGTCACTTGATGACGAATGACGCC
ATCGTCGACACTGACGGCAGCGAGGTGTTGAAAGGCATCATGGATGCCCTATTCAACGGCCTGATC
GCCATCCACGGGCTAAAGGCCAGCGACGTCAACGGGCCGTGATCAACAGCCGACCCGGCTCCATC
TACATCGTCAAGCCGAAGATGCACGGTCCGGCCGAGGTGGCGTTACCTGCGAACTGTTCAGCGG
GTTGAAGATGTGCTGGGTTGCCGAAAACACCATGAAGATCGGCATCATGGACGAGGAACGCCGG
ACCACGGTCAACCTCAAGGCGTGCATCAAAGCTGCCGGACCGCGTGGTGTTCATCAACACCGGG
TTCTGGACCGCACCGCGATGAAATCCACACCTCGATGGAGGCCGGCGATGGTGGCAAGGGC
ACCATGAAGAGCCAGCCGTGGATCTGGCTACGAGGACCACACGTCGATGCCGGCTGGCCGCC
GGGTTCAGCGGCCGAGCCCAGGTGGCAAGGGCATGTGGACAATGACCGAGCTGATGCCGCACATG
GTCGAGACAAAATCGCCCAGCCGCGCCGGGCCAGCACCGCCTGGTTCCTCTCCACTGCG
GCCACCCCTGCATGCGCTGCACTACCACCAAGGTGACGTGCCGCCGTGCAACAAGGACTGGCGGG
AAGCGTCGCCACCATGAAACAATTGCTGACCATTCCGCTGGCAAGGAATTGGCCTGGCTCCC
GACGAGATCCGCGAAGAGGTGACAAACAACGTCAATCCATCCTCGGCTACGTGGTTCGCTGGGTT
GATCAAGGTGTCGGCTGCTCGAAGGTGCCGACATCCACGACGTGCGCTATGGAGGACCGGGCC
ACGCTGCAATCTCAGCCAATTGTTGCCAACTGGCTGCCACGGTGTGATCACCAGCGCGGAT
GTGCGGGCCAGCTTGGAGCGGATGGGCCGTTGGTGTGATCGACAAAACGCCGGGCGACGTGGCATA
CGACCGATGGCACCCAACCTCGACGACAGTATCGCTTCTGGCCGCGCAGGAGCTGATCTGTCC
GGGCCAGCAGCCAACGGTACACCGAGCCATCTGCAACGTCGTCGGAGTTAACGCC
CGGGCCGCTGAGAACGCCGGCCCATCGGACAGGGCCGGTGACGATGCCGCCAGGGTGCAGAAGTAC
GGCGGATCTCGGTTGGCCGACGCCAACGGATTGCGCGTGCACGCACTGTCGCCACCAAG
AAGCAAGGCAATGACGTCGTCGTCGTCGCTCTGCCATGGGGATACCACCGACGACCTGCTGGAT
CTGGCTCAGCAGGTGTGCCCGCGCCCTCGGGAGCTGGACATGCTGCTTACGCCGGTGAA
CGCATCTCGAATGCGTTGGGCCATGGCATCGAGTCGCTGCCGCGCATGCCGGTGTGTCACC
GGTCGCAAGGCCGGGTGATCACCAACCGGACCCACGGCAACGCCAAGATCATCGACGTACGCCG
GGCGGCTGCAAACGCCCTTGAGGAGGGCGGGTCGTTTGCTGGCCGGATTCCAAGGGTCAGC
CAGGACACCAAGGATGTCACGACGTTGGCCGCGGCTCGGACACCACCGCGTGCCTGCCATGCC
GCCGCGCTGGGTGCCGATGTTGAGGAAATGCTGAGATCTACACCGACGTGGACGGCATCTTCAGGCCGACCCG
CGCATCGTGCACGCCGAAAGCTGACACCGTGACCTCGAGGAAATGCTGAGATGGCGGCC
TGCAGGCCAAGGTGCTGATGCTGCGCTGCGTGGAAATACGCTGCCGCCATAATATTCCGGTGCAC
GTCCGGTCTCGTACTCGGACAGACCGGGACCGTCGTTGCGGATCGATCAAGGACGTACCCATG

Figure 1: Nucleotide Sequence of TbF14
Sheet 4 of 4

GAAGACCCCCATCTGACCGGAGTCGCACGACCGCAGCGAGGCCAAGGTGACCATCGTGGCTG
CCCGACATCCCCGGTATGCGGCCAAGGTGTTAGGGCGTGCCAGACGCCACGTAAACATCGA
CATGGTGCTGCAGAACGCTCCAAGGTCGAGGACGGCAAGACCGACATCACCTCACCTGCTCCCG
CAGACGTCGGGCCGCCGTGGAAAAACTGGACTCGCTCAGAACGAGATCGGCTCTACACAG
CTGCTGTACGACGACCACATCGCAAGGTATCGCTGATCGGTGCCGGCATGCGCAGCCACCCGGG
GTCACCGCGACGTTCTGTGAGGCCGTGGCGCGTGGGGTCAACATCGAGCTGATCTCACCTCG
GAAGATCAGAGATCTCGGTGTTGTGCCGCACACCGAAGTGGACAAGGCCGTGGCGCTGCATG
AAGCCTTCGGCTCGGCCGACGAGGAGGCCACGGTGTACGCCGGACGGGACGGTAGATGGGCC
TGTCAATAGTGAATTCATCGATGTGCAGATATCCATCACACTGGCGCCGCTCGAGCACCACCA
ACCACCACTGAGATCCGGCTGCTAACAAAGCCCAGAAGGAAGCTGAGTTGGCTGCTGCCACCGCTG
AGCAATAACTAGCATAACCCCTGGGCCTCTAAACGGGTCTTGAGGGGTTTTGCTGAAAGGAG
GAACTATATCCGGAT

Figure 2: Nucleotide sequence of TbF15
Sheet 1 of 4

FEATURES	Location/Qualifiers
misc_feature	5072..5095
	/note="His tag coding region"
misc_feature	5096..5293
	/note="Ra3 coding region"
misc_feature	5294..6346
	/note="38kD coding region"
misc_feature	6347..6643
	/note="38-1 coding region"
misc_feature	6644..8023
	/note="FL TbH4 coding region"

TGGCGAATGGGACGCGCCCTGTAGCGCGCATTAAGCGCGCGGGTGTGGTGGTTACGCGCAGCGT
GACCGCTACACTGCCAGCGCCCTAGCGCCCGCTCTTCGCTTCTCCCTTCCTTCGCCCC
GTTCGCCGGCTTCCCGTCAAGCTCTAAATCGGGGGCTCCCTTAGGGTCCGATTAGTGCCTT
ACGGCACCTCGACCCCCAAAAACTTGATTAGGGTGTGGACTCCACGTTCTTAATAGTGGACTCTTGTCCAAACTGG
GACGGTTTTCGCCCTTGACGTTGGAGTCCACGTTCTTAATAGTGGACTCTTGTCCAAACTGG
AACAAACACTCAACCTATCTCGGTCTATTCTTTGATTATAAGGGATTTGCCGATTCGGCCTA
TTGGTTAAAAATGAGCTGATTAACAAAATTAAACGGAATTTAACAAAATTAAACGTTAC
AATTCAGGTGGCACTTTCGGGAAATGTGCGCGAACCCCTATTGTTATTCTCAAATACA
TTCAAATATGTATCCGCTCATGAATTAAATTCTTAGAAAACATCGAGCATCAAATGAAACTGCA
ATTATTATCATATCAGGATTATCAATACCATTGGAAAGCCGTTCTGTAATGAAGGAGAAA
ACTCACCGAGGCAGTCCATAGGATGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCAA
CATCAATACAACCTATTAAATTCCCGTCAAAATAAGTTATCAAGTGAGAAATCACCAG
TGACGACTGAATCCGTGAGAATGGAAAAGTTATGCATTCTTCCAGACTTGTCAACAGGCC
AGCCATTACGCTCGTCATCAAATCACTCGCATCAACCAAACCGTTATTCTCGTATTGCC
GAGCGAGACGAAATACGCGATCGTGTAAAGGACAATTACAAACAGGAATGCAACCGGC
GCAGGAACACTGCCAGCGCATCAAATATTTCACCTGAATCAGGATATTCTCTAACACTGG
ATGCTGTTTCCCAGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCT
TGATGGTCGGAAGAGGCATAAATTCCGTAGCCAGTTAGTCTGACCATCTGTAAACATCAT
TGGCAACGCTACCTTGCCATGTTCAGAAACAACCTCTGGCGCATGGCTTCCATACAATCGAT
AGATTGTCGACCTGATTGCCGACATTATCGCGAGCCATTATACCCATATAAACGATCCA
TGTTGGAATTAAATCGCGCCTAGAGCAAGACGTTCCCGTTGAATATGGCTCATAACACCCCTG
TATTACTGTTATGTAAGCAGACAGTTTATTGTTCATGACCAAAATCCCTAACGTGAGTTTCG
TTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTGAGATCCTTTCTCGCG
GTAATCTGCTGCTGCAAACAAAAACCCCGTACCGCTACAGCGTGGTTGTTGCCGATCAAGAG
CTACCAACTCTTCCGAAGGTAACGGCTTCAGCAGAGCGCAGATAACAAACTGTCCTCTA
GTGTAGCCGTAGTTAGGCCACCACTCAAGAACTCTGTAGCACCCTACATACCTCGCTCTGCTA
ATCCTGTTACCACTGGCTGCGCAGTGGCGATAAGTCGTGCTTACCGGGTGGACTCAAGACGA
TAGTTACCGGATAAGCGCAGCGTGGCTGAACGGGGTTCGTCACACAGCCCAGCTTGGAG
CGAACGACCTACACCGAACTGAGATAACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTCCGAA
GGGAGAAAGCGGGACAGGTATCCGTAAGCGGCAGGGTGGAAACAGGGAGCGCACGAGGGAGCTT
CCAGGGGAAACGCCCTGGTATCTTATAGTCCTGTCGGGTTGCCACCTCTGACTTGAGCGTCGA
TTTTGTGATGCTCGTCAGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGCCCTTTACGG

Figure 2: Nucleotide sequence of TbF15
Sheet 2 of 4

TTCCCTGGCCTTTGCTGGCCTTGCTCACATGTTCTTCCTGCCTTACCCCTGATTCTGTGGAT
AACCGTATTACCGCCTTGAGTGAGCTGATACCGCTCGCCGAGCCGAACGACCGAGCGCAGCGAG
TCAGTGAGCGAGGAAGCGGAAGAGCGCTGATGCCGTATTTCTCCTACGCATCTGTGCCGTATT
TCACACCGCATATACTGGTCACACTCTCAGTACAATCTGCTCTGATGCCGTAGTTAAAGCCAGTATA
CACTCCGCTATCGTACGTGACTGGTCATGGCTGCCGACACCCGCCAACACCCGCTGACGC
GCCCTGACGGGCTTGTCTGCTCCCAGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGAGCTG
CATGTGTAGAGGTTTACCGTCATACCGAAACCGCGAGGCAGCTGCCGTAAAGCTCATCAGC
GTGGTCGTGAAGCGATTACAGATGTCCTGTTCATCCGCTCCAGCTCGTTGAGTTCTCCAG
AAGCGTTAATGTCTGGCTTCTGATAAAAGCGGGCATGTTAAGGGCGTTTTCTGTTGGTCAC
TGATGCCCTCGTGAAGGGGATTCTGTTATGGGGTAATGATAACCGATGAAACGAGAGAGGAT
GCTCACGATACGGTTACTGATGATGAACATGCCGGTTACTGGAACGTTGTGAGGGTAAACA
GGCGGTATGGATGCCGGGACCAGAGAAAAACTCACTCAGGGTCAATGCCAGCGCTCGTTAATAC
AGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAAGATCCGGAACATAATGGTCA
GGCGCTGACTCCCGTTCCAGACTTACGAAACACGGAAACCGAAGACCATTGATGTTGTC
TCAGGTGCGAGACGTTTGCAAGCAGCAGTCGCTCACGTTGCTCGTATCGGTGATTCAATTG
CTAACCAAGTAAGGCAACCCCGCCAGCCTAGCCGGTCTCAACGACAGGAGCAGCATGCGCAC
CCGTGGGGCCGCAATGCCGGATAATGGCTGCTCTGCCGAAACGTTGGTGGCGGGACCA
GACGAAGGCTTGAGCGAGGGCGTCAAGATTCCAATACCGCAAGCGACAGGCCGATCATGTC
GCTCCAGCGAAAGCGGTCTGCCGAAAATGACCCAGAGCGCTGCCGACCTGTCCTACGAGTT
CATGATAAAAGAACAGTCATAAGTGCAGCAGCATAGTCATGCCCGGCCACCGGAAGGAGCT
GACTGGGTGAAGGCTCTCAAGGCATCGTCGAGATCCCGTGCCTAATGAGTGAGCTAACTTAC
ATTAATTGCGTTGCGCTCACTGCCGTTCCAGTCGGAAACCTGTCGCCCCAGCTGCAATTAG
AAATCGGCCAACGCGCGGGAGAGGCGGTTGCGTATTGGCGCCAGGGTGGTTTCTTACCCA
GTGAGACGGCAACAGCTGATTGCCCTCACCGCCTGGCCCTGAGAGAGTTGCAAGCAAGCGTCCA
CGCTGGTTGCCCAAGCAGGGAAAATCCTGTTGATGGTGGTTAACGGCGGGATATAACATGAGC
TGTCTCGGTATCGTGTATCCACTACCGAGATATCCGACCAACCGCAGCCGGACTCGGTAA
TGGCGCGCATTGCCAGCGCCATCTGATCGTGGCAACCAGCATCGCAGTGGAAAGATGCC
CATTCAAGCATTGCAATGGTTGTTGAAAACCGACATGGCACTCCAGTCGCTTCCCGTCTCG
TCGGCTGAATTGATTGCGAGTGAGATATTATGCCAGCCAGCCAGACGCGAGCGCCGAGACAG
AACTTAATGGCCGCTAACAGCGCAGATTGCTGGTACCCAATGCGACCAAGATGCTCCACGCC
GTCGCGTACCGTCTCATGGAGAAAATAACTGTTGATGGGTGTCGGTCAGAGACATCAAGAA
ATAACGCGGAACATTAGTCAGGCAGCTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGT
TAATGATCAGCCCAGTCACCGCGTTGCCGAGAAGATTGTCACCGCCGTTACAGGCTTCGACGC
CGCTCGTCTACCATGACACCACCGCTGGCACCCAGTTGATCGCGAGATTAAATGCC
CGACAATTGCAAGGGCGCGTGCAGGGCCAGACTGGAGGTGGCAAGCCAATCAGAACGACTGTT
TGGCCGCCAGTTGTTGTCGCCAGCGGGTGGGAATGTAATTCACTGCCGCTTACGCC
TTTCCCGCTTTGCCAGAAACGTCGGCTGGCCTGGTTACCAACGCCAACGGTCTGATAAGAGA
CACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTACATTCAACGCC
CTTCCCGGCCATCATGCCATACCGCGAAAGGTTTGCAGGCCATTGATGGTGTCCGGGATCTCGA
CGCTCTCCCTTATGCGACTCTGCATTAGGAAGCAGCCAGTAGTAGGTTGAGGCC
GCCGCCAGGAATGGTGCAAGGAGATGGCGCCAAACAGTCCCCGCCACGGGGCTGCC
ACCATACCCACGCCAACAGCGCTCATGAGGCCGAAGTGGCGAGCCGATCTCCCCATCGTG
ATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTCGGCGCGTGA
GCCGCCACGATGCGTCCG

Figure 2: Nucleotide sequence of TbF15
Sheet 3 of 4

GCGTAGAGGATCGAGATCTCGATCCCGCAAATTAAACGACTCACTATAGGGGAATTGTGAGCGG
ATAACAATTCCCCTCTAGAAATAATTGTTAACCTTAAGAAGGGAGATATACATATGGGCCATCA
TCATCATCATCACGTGATCGACATCATCGGGACCAGCCCCACATCCTGGAACAGGCAGGCGGCGGA
GGCGTCCAGCGGGCGGGATAGCGTCGATGACATCCGCGTCGCTCGGTCAAGCAGGACAT
GGCGTGGACAGCGCCGGCAAGATCACCTACCGCATCAAGCTCGAAGTGTGTTCAAGATGAGGCC
GGCGCAACCGAGGTGTGGCTCGAAACCACCGAGCGGTTCGCCTGAAACGGGCAGGCCGGTAC
TGTGCGACTACCCCCGCGTCGCGGTGACGTTGGGGAGACCGTAGCAGCTGCTCACCC
GCTGTTCAACCTGTGGGTCCGGCTTCACGAGAGGTATCCGAACGTACGATCACCGCTCAGGG
CACCGGTTCTGGTGCCGGATCGCGCAGGCCGCCGGACGGTCAACATTGGGCCTCCGACGC
CTATCTGTCGGAAGGTGATATGGCGCGCACAAGGGCTGATGAACATCGCGTAGCCATCTCCGC
TCAGCAGGTCAACTACAACCTGCCGGAGTGAGCGAGCACCTCAAGCTGAACGGAAAAGTCCCTGGC
GCCCATGTACCAGGGCACCATCAAAACCTGGGACGACCCGAGATCGCTGCGCTCAACCCGGCGT
GAACCTGCCGGCACCGCGGTAGTTCCGCTGCACCGCTCCGACGGTCCGGTACACCTTCTTGT
CACCCAGTACCTGTCGAAGCAAGATCCCAGGGCTGGGGCAAGTCGCCCGCTCGGCACCACCGT
CGACTTCCCGCGGTGCCGGTGCCTGGGTGAGAACGGCAACGGCGATGGTGACCGGGTTGCGC
CGAGACACCGGGCTCGCTGGCTATATCGGCATCAGCTTCCTCGACCAGGCCAGTCAACGGGACT
CGGGGAGGGCCAACTAGGAATAGCTCTGGCAATTCTTGTGCCCCGACGGCAAAGCATTAGGC
CGCGCGGGCTGGCTCGCATCGAAAACCCGGGAACCAGGGGATTCGATGATCGACGGGCCCGC
CCCGGACGGCTACCCGATCATCAACTACGAGTACGCCATCGTCAACAACCGGAAAAGGACGCCGC
CACCGCGCAGACCTTGCAGGCATTCTGCACTGGCGATCACCGACGGCAACAAGGCCTCGTTCC
CGACCAGGTTCATTCCAGCCGTGCCGCCGGTGGTGAAGTTGTCTGACGCGTTGATCGC
GATTTCAGCGCTGAGATGAAGACCGATGCCGCTACCCCTCGCGCAGGAGGAGGTAAATTGAGCG
GATCTCCGGCACCTGAAAACCCAGATCGACCGAGGTGGAGTCGACGGCAGGTTCGTCAGGGCCA
GTGGCGCGCGCGGGGACGGCCGCCAGGCCGGTGGTGCCTCCAAGAACGGCAAA
GCAGAACAGGAACCTGACGAGATCTGACGAATATTGTCAGGCCGGTCCAATACTCGAGGGC
CGACGAGGAGCAGCAGCAGGCCGTGTCCTCGCAAATGGCTTACTCAGTCGACGGTACGGT
GGATCAGCAAGAGATTGACAGGGCAACGAGGTGGAGGCCGATGGCGAACCCACCGACTGA
TGTCCCCATCACACCGTGCAGACTCACGGCGCTAAAAACGCCGCCAACAGCTGGTATTGTC
CGACAACATGCGGAATACCTGGCGGCCGGTGCCAAAGAGCGGAGCGTCTGGCGACCTCGCTGCG
CAACGCCAGGCGTATGGCGAGGTGATGAGGAGGCTGCACCGCGCTGGACAACGACGCCGA
AGGAACCTGTCAGGCAGAACATGCCGGGGCGTCGGAGGGGACAGTTGCCGAACTAACCGATA
GCCGAGGGTGGCCACGCCGGTGAACCCAACCTCATGGATCTCAAAGAACGGCAAGGAAGCTGA
AACGGCGACCAAGGCCATCGCTCGCAGCTTGCAGGATGGGTGGAACACTTCAACCTGACGCT
GCAAGGCCAGTCAAGCGTTCCGGGGTTGACAACCTGGGAGGCCGATGCCGCTACCGCTTGC
GGCTCGCTCGATCAACAACGGCAATGGATACTCCACATGGCAAATTGAGCGCTGC
GCAGGCTCAATATGTCGCGCAGCTGCACGTGTCGGCTAGGCCAACATCCGACTATGAAGACAT
AGTCGGGCTCGAACGGCTTACGCCGGAAAACCCCTGCCGGCGACCAAATTCTCCGGTACGC
GGAGTATCAGCAGAGGTGGAGAAGGTGCTGACCGAATACAACAAGGCAGGCC
AAACCCGCCAGCCTCCCCCGCCATCAAGATCGACCCGCCCGCTCCGCAAGAGCAGGGATT
GATCCCTGGCTTCTGATGCCGCCGTCGACGGCTCCGGTGTGACTCCGGTACCGGGATGCC
CGCACCGATGGTTCCGCCAACCGGATGCCGGGGTGGTGGCCTCCGGCTGACACGGCGCG
GACGTCGGCTGGCGGGAAAGCCGAGCGCTGTCGGCGACGTGGCGGTCAAAGCGGC
TGGCGGTGGAGGCAGGCCGTCGGCGCCGTTGGGATCCGCGATCGGGGGCGCGAATCGGT

Figure 2: Nucleotide sequence of TbF15
Sheet 4 of 4

GCGGCCGCTGGCGTGGT GACATTGCCGGCTTAGGCCAGGAAAGGGCCGGCGGCCGCGCT
GGCGCGCGGTGGCATGGGAATGCCGATGGGTGCCGCGCATCAGGGACAAGGGGGCGCCAAGTC
GGGTTCTCAGCAGGAAGACGAGGCCTCTACACCGAGGATCGGGCATGGACCAGGCCGTATTGG
TAACCGTCGGCGCCAGGACAGTAAGGAGTCGAAGTGAATTCTGCAGATATCCATCACACTGGCGGC
CGCTCGAGCACCAACCACCACTGAGATCCGGCTGCTAACAAAGCCGAAAGGAAGCTGAGT
TGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGCCTCTAAACGGGTCTTGAGGG
GTTTTTGCTGAAAGGAGGAACATATCCGGAT

Figure 3: Amino Acid Sequence of TbF14

MQHHHHHHHTDRSVGNLRIARVLYDFVNNEALPGTDIDPDSFWAGVDKVADLTPQNQALLNARDE
LQAQIDKWHRRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITSGVDAEITTTAGPQLVVPVLNAR
FALNAANARWGSLYDALYGTDVIPETDGAEKGPTYNKVRGDKVIAYARKFLDDSVPLSSGSFGDAT
GFTVQDGQLVVALPDKSTGLANPGQFAGYTGAAESPTSVLLINHGLHIEILIDPESQVGTTDRAGV
KDViLESAITTIMDFEDSVAAVDAADKVLGYRNWLGLNKGDLAAAVDKDTAFLRVLNRDRNYTAP
GGGQFTLPGRSLMFVRNVGHLMTNDAIVDTDGEVFEGIMDALFTGLIAIHGLKASDVNGPLINSR
TGSIIYIVKPKMHGPAAEVAFTCELFSRVEDVLGLPQNTMKIGIMDEERRRTVNLKACIKAADRVVF
INTGFLDRTGDEIHTSMEAGPMVRKGTMKSQPWIAYEDHNDAGLAAGFSGRAQVGKGMWTMTEL
MADMVETKIAQPRAGASTAWVSPTAATLHALHYHQDVAAVQQLAGKRATIEQLLTIPAKEL
AWAPDEIREEVDDNNCQSILGYVWRWDQGVGCKVPDIHDVALMEDRATLRISSQLLANWLRHGVI
TSADVRASLERMAPLVDRQNAGDVAYRPMAPNFDDSIIFLAAQELILSGAQQPNGYTEPILHRRRR
EFKARAEEKPAPSDRAGDDAARVQKYGGSSVADAERIRRAERIVATKKQGNDVVVVVSAMGDTTD
DLDDLAQQVCPAPPRELDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTGTHGNAKII
DVTGRLQTALEEGRVVLVAGFQGVSQDTKDVTLGRGGSDDTAVAMAAALGADVCEIYTDVDGIF
SADPRIVRNARKLDTVTFEEMLEAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGBTVVGSIK
DVPMEDPILTGVAHDRSEAKVTIVGLPDIPGYAAKVFRavarrrrqhrhGAAERLQGRGRQDRHHL
HLLPQTSGPPPWNTRSETRSASTQLYDDHIGKVSLIGAGMRSHPGVTATFCEALAAVGVNIEL
ISTSEDQRSRCCAATPNWTRPWSRCMKRSGAATRRPRCTRGRDGRWACQ..

Figure 4: Amino Acid Sequence of TbF15

MGHHHHHHVIDIIGTSPTSWEQAAAEEAVQRARDSVDDIRVARVIEQDMAVDSAGKITYRIKLEVSF
KMRPAQPRCGSKPPSGSPETGAGAGTVATTPASSPVTLAETGSTLLYPLFNLWGPAGHERYPNVTI
TAQGTGSGAGIAQAAAGTVNIGASDAYLSEGDMAAHKGLMNIALAISQQVNLYNLPGVSEHLKLNG
KVLAAMYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRSRGSGDTFLFTQYLSKQDPEGWGKSPGF
GTVDFPAVPGALGENGNMGMTGCAETPGCVAYIGISFLDQASRGLGEAQLGNSSGNFLLPDAQ
SIQAAAAGFASKTPANQAIMIDGPAPDGYPINYEYAIIVNNRQKDAATAQTLQAFLHWAITDGNK
ASFLDQVHFQPLPPAVVKLSDALIATISSAEMKTDAAATLAQEAGNFERISGDLKTQIDQVESTAGS
LQGQWRGAAGTAAQAAVVRFQEAAKQKQELDEISTNIHQAGVQYSRADEEQQQALSSQMGTQSQ
TVTVDQQEIILNRANEVEAPMADPPTDVPITPCELTAAKNAQQQLVLSADNMREYLAAGAKERQRLA
TSLRNAAKAYGEVDEEAATALDNDGEHTVQAESAGAVGGDSSAELTDTPRVATAGEPNFMDLKEAA
RKLETGDQGASLAHFADGWNTFNLTLQGDVKRFRGFDNWEGDAATACEASLDQQRQWILHMAKLSA
AMAKQAQYVAQLHVWARREHPTYEDIVGLERLYAENPSARDQILPVYAEYQQRSEKVLTEYNKAA
LEPVNPPKPPPAAIKIDPPPPQEQQLIPGFLMPPSDGSGVTPGTGMPAAPMVPPGTGSPGGGLPADT
AAQLTSAGREAAALSGDVAVKAASLGGGGGGVPsapLGSIAIGGAESVRPAGAGDIAGLGQGRAGG
GAALGGGGMGMPGAAHQGQGGAKSKGSQQEDEALYTEDRAWTEAVIGNRRQDSKESK.

Figure 5

	Status	TbF15	TbF6
5004	TB	0.926	1.045
7004	TB	0.928	1.184
9004	TB	1.102	1.365
11004	TB	0.856	1.629
15004	TB	2.035	2.099
17004	TB	2.893	2.867
18004	TB	0.477	0.414
21004	TB	1.062	1.635
23004	TB	0.429	0.501
26004	TB	0.299	0.392
27004	TB	0.244	0.207
28004	TB	2.236	2.04
30004	TB	2.058	1.508
32004	TB	2.324	1.927
33004	TB	1.600	1.578
34004	TB	1.059	1.136
36004	TB	0.546	1.105
37004	TB	1.446	1.989
39004	TB	2.021	2.782
41004	TB	0.511	0.652
43004	TB	0.855	0.483
44004	TB	0.731	0.66
53004	TB	1.100	0.317
FD8-24	Control	0.183	0.314
FD8-25	Control	0.061	0.063
FD8-26	Control	0.066	0.142
FD8-27	Control	0.021	0.115
FD8-28	Control	0.053	0.289
FD8-29	Control	0.114	0.238
FD8-30	Control	0.105	0.146
FD8-31	Control	0.101	0.237
FD8-33	Control	0.080	0.071
FD8-34	Control	0.140	0.117
FD8-35	Control	0.088	0.072
FD8-36	Control	0.081	0.089
FD8-37	Control	0.057	0.06
FD8-38	Control	0.104	0.111
FD8-39	Control	0.221	0.241
FD8-40	Control	0.257	0.265
FD8-41	Control	0.056	0.093
FD8-42	Control	0.184	0.273
FD8-43	Control	0.126	0.126
FD8-44	Control	0.193	0.092
FD8-45	Control	0.058	0.057
FD8-46	Control	0.183	0.23
FD8-48	Control	0.062	0.085
FD8-49	Control	0.134	0.247
Mean		0.113	0.157
SD		0.061	0.086
Mean +3SD		0.298	0.414
Sensitivity		22/23	20/23

Monday, July 26, 1999 10:42 AM

HTCC-1.seq.mpd, (1 > 1200) Site and Sequence

Enzymes: All 515 enzymes (No Filter)

Settings: Circular, Certain Sites Only, Standard Genetic Code

HTCC-1 FL Sequence

Page 1

CAGGCATGAGCAGAGCGTTCATCACTCGATCCAACGATCAGTGGCATTGACGGCTTGACGACCTTGAGGAATGGAAATACCCAAACCAAGGGGTATGCT

GTCGGTACTCGTCTCGGAAGTAGTAGCTAGGTTGCTAGTCACGGTAACCTGGAAACATGCTGGAAAGACCCCTAACCTTATGGTTGGTCCCCC4TAGGA + 100

HTCC-1 FL

M S R A F I I D P T I S A I O G L Y O L L G I G I P N O G G I L

TTACTCCCTCACTAGAGTACTTCGAAAAAGCCCTGGAGGAGCTGGCAGCGCTTCCGGGTGATGGCTGGTTAGGTTGGCCCGGGACAAATACGCCGGC + 200

AATGAGGAGTGATCTCATGAAGCTTTTCGGGACCTCGACCGTCTCGACCGCAACTACCGGACCAATCCAAAGCCGGCCCTGTTATGGGGCG + 200

HTCC-1 FL

Y S S L E Y F E K A L E E L A A A F P G D G W L G S A A O K Y A G

AAAAACCGCAACCACCGTGAATTTCAGGAACCTGGCAGACCTCGATCGTCAGCTCATCAGCCTGATCCACGGACAGGCCAACCGGGTCCAGACGACCC + 300

TTTTGGCGTTGGTCACTTAAAGGTCTTGACCGTCTGGAGCTAGCAGTCGAGTAGTCGGACTAGGTGGTGTGGTTGCGCCAGGTCTGCTGGG + 300

HTCC-1 FL

K N R N H V N F F Q E L A O L D O R Q L I S L I H D Q A N A V C T T

GCGACATCCCTGGAGGGCGCCAAAGAAAGGTCAGGTTCTGCGGCGCCGGTGGCTGTGGACCTGACCTACATCCGGTCTGGCCACGCCATCGGCCG + 400

CGCTGTAGGACCTCCCGCGGTTCTTCAGAGCTAACGACGGGGCACCGACACCTGGACTGGATGTAGGGCCAGCAGCCGTGCGGGATAGCCGGCG + 400

HTCC-1 FL

R D I L E G A K K G L E F V R P V A V O L T Y I P V V G H A L S A A

CTTCAGGCGCCGTTTGCAGGGCGCGATGGCCGTAGTGGCGGGCGCTTGCCTACTTGGCTGTAACCGCAGCTAACCTCCCTCAAACCGCAGCTCAAC + 500

GAAGGTCCCGGGAAAACCGGCCCGCGTACCGGCATCACCCGCCCGGAACGGATGAACCAGCACTTTGGACTAGTTGCCTGAGTTGAGGAGTT + 500

HTCC-1 FL

F Q A P F C A G A M A V V G G A L A Y L V V K T L I N A T Q L L K

TTGCTTGCCAAATTGGCGGAGTTGGTCGCGCCGCAATTGGACATCATTCGGATGTGGCGACATCATCAAGGGCACCCCTGGAGAAGTGTGGAGT + 600

AACGAACGGTTAACCGCCTAACCCAGCGCGGGTAACGGCTGTAGTAAAGCCTACACCCCTGTAGTAGTTCCCGTGGAGCCCTTCAACCCCTCA + 600

HTCC-1 FL

L L A K L A E L V A A A I A O I I S O V A D I I K G T L G E V W E

TCATCACAAACCGCGCTAACGGCCTGAAAGAGCTTGGACAAGCTCACGGGGTGGGTGACCGGACTGTTCTCTCGAGGGTGGTCAACCTGGAGTCCTT + 700

AGTAGTGTGCGAGTTGCCGACTTCTCGAAACCCCTGGAGTCCCCACCCACTGGCT3ACAAGAGAGCTCCACCCAGCTGGACCTCAGGAA + 700

HTCC-1 FL

F I T N A L N G L K E L W D K L T G W V T G L F S R G W S N L E S F

CTTGCGGGCGTCCCCGGCTTGACCGGGCGACCCAGCGGCTTGTGGCAAGTGAATGGCTTGTGGTGGCCGCTGTCCGCATCGTCGGCTTGGCT + 800

GAAACGCCGCCAGGGGCCGAACGGCCGCGCTGGTGGCGAACAGCGTTACTGACCGAACAGCCACCCGGCAQACAGCGTAGCAGCCGAACCGA + 800

HTCC-1 FL

F A G V P G L T G A T S G L S Q V T G L F G A A G L S A S S G L A

Fig. 6

sheet 1 of 2

Monday, July 26, 1999 10:42 AM

TCG#1 seq.mod. (1 > 1200) Site and Sequenc:

Page 2

CACGCCGGATAGCCTGGCGAGCTCAGCCAGGTGCCCGCCCTGGCGGGCATGGGGGGGGTCCGGTTTGGGGCTTGCCGAGCTGGCTCAGGTCCATG
GTGCCCTATCGGACCGCTCGAGTCGGTCGAACGGGCGGGACCAGGGTAACCCCCCGCCAGGCCAAAACCCCCCAACGGCTCGGACCGAGTCAGGTAC + 900

HTCC-1 FL

H A D S L A S S A S L P A L A G I G G G S G F G G L P S L A Q V H

CCGCCCTCAACTCGGCAGGCCTACGGCCCGAGCTGATGGCCCGGCTGGCGGCCCTGCCGAGCAAGTCGGCGGGAGTCGAGCTGGTCTCCGCGCAAGGG
GGCGGAGTTGAGCCGTCCGGATGCCGGGGCTCGACTACCGGGCAGCCCGGGQACGGCTCGTCCAGCCCQCCGTCAGCGTCGACCAGAGGCCGTCGCCC + 1000

HTCC-1 FL

A A S T R Q A L R P R A Q G P V G A A A E Q V G G Q S O L V S A Q G

TTCCAAGGTATGGGCGGACCGTAGGCATGGCGGCATGCACCCCTCTTCGGGGCGTCGAAAGGGACQACGAGGAAGAAGTACTCGGAAGGCCGCGCG
AAGGGTTCCATACCCGCCTGGCATCCGTACCGTGGGAGAAGCCCCCGAGCTTCCCTGCTGCTGTTCTCATGAGCCTTCGGCGCCGC + 1100

HTCC-1 FL

S Q G M G G P V G M G G M H P S S G A S K G T T T K K Y S E G A A

GCGGGCACTGAAGACGCCGAGCGCGGCCAGTCGAAGCTGACGGGGGGTGGCAAAAGGTGCTGGTACGAAAGTCGTAAACGGCATGGCGAGCAA
CGCCCGTGACTTCTCGCGCTCGCGCGGTCACTTCGACTGCGCCGCCACCCGTTTCCACGACCATGCTTGCAGCAGATTQCCGTACCGCTCGGTT + 1200

HTCC-1 FL

A G T E D A E R A P V E A O A G G G Q K V L V R N V V

biochemistry lab

FIG. b

Sheet 2 of 2

Monday, July 26, 1999 10:49 AM

TCC1(1-232) Map.mpd (1 > 728) Sita and Seq. sa

Enzymes : 212 of 315 enzymes (Filtered)

Settings: Linear, Certain Sitas Only, Standard Genetic Code

H7CE-1 (1-232)

Page 1

ATGCATCACCATCACCATCACATGAGCAGAGCTTCATCACTCGATCCAACGATCAGTGGCAATTGACGGCTTGACGGCTTGACGGCTTGACGGATTGGAAATACCCA
+ 100
TAGGTAGTGGTAGTGGTAGTGTACTCGCTCGAAGTAGTACGTTAGTCAACGGTAACCGCCGAAACATGCTCGAAGACCCCTAACCTTATGGGT
M H H H H H M S R A F I I O P T I S A I O G L Y D L L G I G I P

ACCAAGGGGGTATCCCTTACTCCCTCACTAGAAGTACTTGAAAGAAAAGCCCTGGAGGAGCTGGCAGCAGCGTTCCGGGTATGGCTGGTAGGTTCGGCG
+ 200
TGGTCCCCCATAGGAAATGAGGAGTGATCTCATGAACCTTTTCGCGACCTCTCGACCCCTCGTCAAAAGGCCAGTACCGACCCAATCCAAGCCGGCG
N Q G G I L Y S S L E Y F E K A L E E L A A A A F P G D G W L G S A A -

GGACAAATAACGCCGGCAAAACCGCAACCACTGAATTTCAGGAACCTGGCAGACCTCGATCGTCAGCTCATCAGCCTGATCCACGACCCAGGCCAAC
+ 300
CCTGTTTATGCGGCCGTTTGGCGTTGGCACTTAAAAAGGTCTTGACCGTCTGGAGCTAGCAGTCAGTAGTGGACTAGGTGCTGGTCCGGTTG
O K Y A G K N R N H V N F F Q E L A D L O R Q L I S L I H O O A N

GCGGTCCAGACGACCCCGCAGATCCTGGAGGGGCCAAGAAAGGTCTCGAGTTCGTGGCGCCCGTGGCTGTGGACCTGACCTACATCCCGTCTGGCG
+ 400
CGCCAGGTCTGCTGGCGCTGTAGGACCTCCCGGGTTTTCCAGAGCTCAAGCACCCGGCCACCGACACCTGGACTGGATGAGGGCCAGCAGCCCG
A V Q T T R O I L E G A K K G L E F V R P V A V O L T Y I P V V G

ACGCCCTATCGGCCGCTTCCAGCGCCGTTTGCAGCGCGAAGGCCGTAGTGGCGCCGCGCTTGCTACTTGGCTGAAAACGCTGATCAACGC
+ 500
TGCGGGATAGCCCCCGGAAGGTCCCGGGCAAAACCGCCCGCGTACCGGCATCACCCCGCCGCAACGGATGAACCCAGCACTTTGCAGTAGTTGCG
H A L S A A F Q A P F C A G A M A V V G G A L A Y L V V K T L I N A

GACTCAACTCCTCAAATTGCTGCCAAATTGGCGAGTTGGTCGCGCCGCTATTGGGACATCATTCGGATGTGGGGACATCATCAAGGGCATCTC
+ 600
CTGAGTTGAGGAGTTAACGAACGGTTAACCGCCTCAACCAGCGCCGGTAACCGCTGTAGTAAAGCCTACACCCCTGTAGTAGTTCCCGTAGGAG
T Q L L K L L A K L A E L V A A A I A D I I S O V A D I I K G I L

GGAGAAAGTGTGGAGTTCATCACAAACCGCCTCAACGGCTGAAAGAGCTTGGGACAAGCTCACGGGTGGGTGACGGACTGTTCTCGAGGGTGGT
+ 700
CCTCTCACACCCCTCAAGTAGTGTGGTCGCGAGTTGCGGACTTCTCGAAACCCCTGTTGAGTGGCCCACCCACTGGCCTGACAAGAGAGCTCCACCA
G E V W E F I T N A L N G L K E L W O K L T G W V T G L F S R G W

CGAACCTGGAGTCCCTTAAGAATT
+ 726
GCTTGGACCTCAGGAAGATTCTTAAG
S N L E S F . E F

FIG. 7a

Monday, July 26, 1999 10:50 AM

HTCC1(184-372) Map.mod (1 > 661) Site and Sequence

Enzymes : 212 of 615 enzymes (Filtered)

Settings: Linear, Certain Sites Only, Standard Genetic Code

HTCC-1 (184-372)

Page 1

ATGCATCACCATCACCATCACCATGTGGCGACATCATCAAGGGCATCCTCGGAGAAGTGTGGAGTTCATCACAAACGGGCTCAACGGGCTGAAAGAGC 100
TACGTACTGGTAGTGTTAGTGTTACACCGCCCTGTAGTAGTTCGGTAGGGAGCCTCTTCACACCCCTCAAGTAGTGTGGCGAGTTGGCGGACTTTCTCG
M H H H H H D V A O I K G I L G E V W E F I T N A L N G L K E
TTTGGGACAAGCTCACGGGGTGGGTGACCGGACTGTTCTCGAGGGTGTGCGAACCTGGAGTCCTTCGGCTGGCGCTCCCCGGCTTGACCGGGCGAC 200
AAACCCGTTCGAGTGCCCCACCCACTGGCCCTGACAAGAGAGCTCCACCAGCTTGGACCTCAGGAAGAAAACCCCGCAGGGGCCGAACCTGGCCCGCTG
L W D K L T G W V T G L F S R G W S N L E S F F A G V P G L T G A T -
CAGCGGCTTGTGCGAACGTGACTGGCTTGTTCGGTGCAGGGCTCTGTCCGCATCGTCGGCTTGGCTCACCGGGATAGGCTGGCGAGCTCAGCCAGCTG 300
GTCGCCGAACAGCGTTCACTGACCGAACAGCCACGCCGGCAGACAGGGTAGCAGCCGAACCGAGTGCCTATCGGACCCQCTCGAGTCGGTCAAC
S G L S O V T G L F G A A G L S A S S G L A H A D S L A S S A S L
CCCCCCTGGCCGGCATTGGGGGGGGTCCGGTTGGGGCTTGCGAGCCCTGGCTCAGGTCCATGCCGCCCTCAACTCGGCAGGGCTACGGGCCCCGAG 400
GGGGGGGACCGGGCGTAACCCCCGCCAGGCCAAAACCCCCGAAAGGCTCGGACCGAGTCCAGGTACGGGGAGTTGAGCCGTCCGGATGCCGGGGCTC
P A L A G I G G S G F G L P S L A O V H A A S T R Q A L R P R
CTGATGGCCCGGTGGCGGCCGTGCCAGCTGGCGGGCAGTCGGAGCTGGTCTCCGCAGGGTTCCCAAGGTATGGCGGACCCGTAGGCATGGG 500
GACTACCGGGCCAGCCGCGGGCAGCCCTCGTCCAGCCCGGTCAAGCTCGACCAGAGGGCGCTCCCAAGGTTCCATACCCGCCCTGGGATCCGTACCC
A D G P V G A A A E Q V G G D S Q L Y S A Q G S D G M G G P V G M G
CGGCATGGACCCCTCTTCGGGGGGCTGAAAGGGACGACGAAGAAGTACTCGGAAGGGCGGGCGGGACTGAAGACGCCAGCGCCAGTC 600
GCCGTACGTGGGGAGAAGCCCCCGAGCTTCCCTGCTGCTTCTTCATGAGGCTTCCGGGCCGCCGACTGACTTCTGCGCTCGCGCGGGTCAG
G M H P S S G A S K G T T T K K Y S E G A A A G T E O A E R A P V
GAACCTGACGGGGCGGTGGCAAAAGGTGGTACGAAACGTGCTTAACGGCGAATTTC 661
CTTCGACTGCGCCGCCACCCGTTTCCACGACCATGCTTGCAGCAGATTGCCGCTTAAG
E A D A G G G Q K V L Y R N V V . R R I ..

FIG. 7b

onday, July 26, 1999 10:46 AM

Ka12-HTCC-1

Page 1

HTCC1.mdf (1 > 1629) Site and Sequence

Enzymes: All STS Enzyme (No Filter)

Settings: Linear, Certain Sites Only, Standard Genetic Code

CATATGCATCACCATCACCATCACACGGGGCGCTCGGATAACTTCCAGCTGTCCCAGGGTGGCAGGGATTGCCATTCCGATCGGGCAGGGATGGCGA

100

GTATACGTTAGTGGTAGTGGTAGTGTGCCGGCGCAGGGATTGAAGGTCGACAGGGTCCCACCCGTCCTAAAGCGGTAAAGGCTAGCCCCTCGCTACCGCT

Met / HIS TAG

Ra12

H M H H H H H T A A S D N F Q L S Q G G Q G F A I P I G Q A M A

TGGCGGGCCAGATCGGATCGGGTGGGGGGTCAACCCACCGCTTCATATCGGGCCTACCGQCTTCTCGGCTTGGTGTGTGACAACAACGGCAACGGCGC

200

ACCCCGGGCTAGGGTAGGCCAACCCCCCAGTQQQTGGCAAGTATAGCCCGATGGCGGAAGGAGCCGAACCCACAACAGCTGTTGCCGTTGCCGCG

Ra12

I A G Q I R S G G S P T V H I G P T A F L G L G V Y V O N N N G N G A

ACGAGTCCAACGGTGGTCGGAGCGCTCCGGCGCAAGTCTGGCATCTCCACCGGCGACGTGATCACCGCGTCGACGGGCTCCGATCAACTCGGCC

300

TGCTCAGGTTGCCACCAGCCCTCGCGAGGGCGCCGTTAGAGCGTAGAGGTGGCCGCTGCACTAGTGGGCCAGCTGCCGAGGGTAGTTGAGCCG

Ra12

R V Q R V V G S A P A A S L G I S T G O V I T A V D G A P I N S A

ACCGCGATGGCGGACCGCGCTTAACGGGATCATCCCGTGACGTATCTGGTGACCTGGCAAACCAAGTGGGGGGCACGGTACAGGGAACGTGACAT

400

TGGCGCTACCGCCTGCCGAATTGCCGTAGTAGGGGACTGCACTAGAGCCACTGGACCGTTGGTTCAAGGCCGCTGCCATGTCCTTGCACTGTA

Ra12

T A M A D A L N G H H P G O V I S V T W Q T K S G G T R T G N V T

TGGCCAGGGACCCCCGGCGAATTCTAGTACCTAGAGGTTCAATGACCGAGAGCGTTCATCATCGATCCAACGATCAGTGGCATTGACGGCTTGACGA

500

ACCGGCTCCCTGGGGCCGCTTAAGGATCATGGATCTCAAGTTACTCGTCTCGCAAGTAGCTAGGTTGCTAGTCACGGTAACGCCAACATGCT

Ra12 EcoRI Thrombin

HTCC1

L A E G P P A E F L V P R G S M S R A F I I D P T I S A T I D G L Y D

CCTTCTGGGATTGGAATACCAACCAAGGGGTATCCTTACTCCTCACTAGAGTACTTCGAAAAAGCCCTGGAGGAGCTGGCAGCGTTCGGGT

600

GGAAGACCCCTAACCTTATGGGTTGGTCCCCATAGGAATGAGGAGTGTCTCATGAAGCTTTTGGGACCTCTCGACCGTCGTCGCAAAGGCGCA

htcc1

L L G I G I P N Q G G I L Y S S L E Y F E K A L E E L A A A F P G

GATGGCTGGTTAGGTTGGCCCGGGACAAATACGGCGGCAAAACCGCAACCGTGAATTTCAGGAACCTGGCAGACCTCGATCGTCAGCTCATCA

700

CTACGGACCAATCCAAGCCGGCGCTGTTATGGGCGTTGGCTGGGACTTAAAGGTCCTTGACCGCTGGAGCTAGCAGTCAGTAGT

htcc1

D G W L G S A A D K Y A G K N R N H V N F F Q E L A D L O R Q L I

GCCTGATCCACGACCAAGGCCAACGGGCTCCAGACGACCCCCGACATCCTGGAGGGCCCCAAGAAAGGTCGAGTTGTGGCCGGCTGGCTGGACCT

800

CGGACTAGGTGCTGGTCCGGTTGGCCAGGTGCTGGCGCTGTAGGACCTCCCGCGGTTCTTCAGAGCTCAAGCACGCCGGCACCGACACCTGG

htcc1

S L I H O Q A N A V Q T T R D I L E G A K K S L E F V R P V A V D L

GACCTACATCCGGTCGTCGGGACGCCATCGGCCCTATCGGCCGCTTCCAGGCCGCTTGGCGGGCGCGATGGCCGTAGTGGCGGGCGCTTGCCCTACTTG

900

CTGGATGTAGGGCCAGCAGCCCGTGCAGGGATAGCCGGCGAAGGTCCGCGCAAAACGCCCGCGCTACCGGCAATCACCCGCCGCGAACGGATGAAC

htcc1

T Y I P V V G H A L S A A F Q A P F C A G A M A V V G G A L A Y L

FIG. 8

Sheet 1 of 2

3(Thr)hTCC1.mod (1 > 1629) Site and Sequence

ATCGTAAAAACGCTGATCACCGCACTCAACTCTCAAATTGGTTGCCAAATTGGCGGAGTTGGTCGGCGCGCCATTGGGACATCATTTGGATGTGG → 1000
 CAGCACTTTGCGACTAGTTGCCCTGAGTTGAGGAGTTAACGAACGGTTAACCGGCCCTAACCGGCCGGCGTAACGGCTGTAGTAAAGCCCTACACC
hTCC1
 V V K T L I N A T G L L K L L A K L A E L V A A A I A D I I S O V
 CGGACATCATCAAGGGCATCCCTCGGAGAAGTGTGGAGTTCATCACAAACGGCCTCAACGGCCTGAAAGAGCTTGGGACAAGCTCACGGGTGGGTGAC → 1100
 GCCTGTAGTAGTTCCCGTAGGGGCCCTTCACACCGCTCAAGTAGTGTTGGCGAGTTGCCGGACTTTCTCGAAACCGCTGGAGTGGCCACCCACTG
hTCC1
 A O I I K G I L G E V W E F I T N A L N G L K E L W O X L T G W V T
 CGGACTGTTCTCGAGGGTGGTCGAACCTGGAGTCCTTCTTCCGGGCGTCCCCGGCTTGACCGGGCGCAGCCAGGGCTTGTGCGCAAGTQACTGGCTTG → 1200
 GCCTGACAAGAGAGCTCCACCCAGCTGGACCTCAGGAAGAAAAGGCCCGCAGGGGCCGAACCTGGCCCGCTGGTCGGCAACAGCGTTACTGACCGAAC
hTCC1
 G L F S R G W S N L E S F F A G V P G L T G A T S G L S Q V T G L
 TTGGTGCAGGCCGGCTGTCCGCATCGCGGCTTGGCTCACGGGATAAGCCTGGCGAGCTAGCCAGCTTGCCCCGGCTGGCGGCGATTGGGGGGGGT → 1300
 AAGCCACGCCGGCCAGACAGGGCTAGCAGCCGAACCGAGTQCGCCTATCGGACCGCTCGAGTCGGTCAACGGGGGGACCGGGCGTAACCCCGCCCA
hTCC1
 S F G A A G L S A S S G L A H A D S L A S S A S L P A L A G I G G G
 CGGGTTTTGGGGCTTGGCGAGCCCTGGCTCAGGTCATGCCCGCTCAACTCGGAGGGCTACGGCCCCGAGCTGATGGCCGGCTGGCGCGCTGGCGA → 1400
 GGCCAAAACCCCCGAAACGGCTGGACCGAGTCCAGGTACGGGGAGTTGAGCCGTCCGCGATGCCGGGCTCGACTACCGGGCCAGCCCGGGCGACGGCT
hTCC1
 S G F G G L P S L A Q V H A A S T R Q A L R P R A D G P V G A A A E
 GCAGGTGGCGGGCAGTCGCAGCTGGCTCCGCGCAGGGTCCAAAGGTATGGCGGACCCCTAGGCATGGCCGGCATGCACCCCTTCGGGGCGTCG → 1500
 CGTCCAGCCGCCGTAGCGTCGACCAAGGGCGCTCCAAAGGGTTCCATACCGCCCTGGCATCCGTACCCGGCTACGTGGGAGAAGCCCCCGCAGC
hTCC1
 Q V G G Q S Q L V S A Q G S Q G M G G P V G M G G M H P S S G A S
 AAAGGGACGACGACGAAGAAGTACTCGGAAGGCAGGGGGCACTGAAGACGCCAGGCCGCGCAGTCGAAGCTGACQGGGGGGTGGGAAAGG → 1600
 TTTCCCTGCTGCTGCTTCTCATGAGCCTTCCGCGCCGCCCGTGAECTCTGCGGCTCGCGCGCGTCAAGCTTCQACTGCGCCGCCACCCGTTTCC
hTCC1
 K G T T T K K Y S E G A A A G T E D A E R A P V E A D A G G G Q K
 TGCTGGTACGAAACGTCGTCTAAGAATTTC → 1629
 ACGACCATGCTTGCAGCAGATTCTAAG
hTCC1 EcoRI
 V L V R N V V . E F

FIG. 8

Sheet 2 of 2

hursday, July 22, 1999 1:35 PM

TCC1(-TD.1).mod (1 > 1225) Site and Sequen-

enzymes : 2 of 515 enzymes (Filtered)

Settings: Linear, Certain Sites Only, Standard Genetic Code

HTCC-1 (T27.1)

Page 1

CATATGCATCACCATCACCATCACATGAGCAGAGCTTCATCATCGATCCAAACGATCACTGCCATTGACGGCTGTGACGGACTCTGGGATTTGGAATAC
100
GTATACTGAGTGGTAGTGGTAGTGTACTCGTCTCGAAAGTAGTAGCTAGGTTGCTAGTCACGGTAACCTGCCAACATCTGCAAGACCCCCTAACCTTATG
Met/HIS TAG hTCC1

H M H H H H H M S R A F I I O P T I S A I D G L Y D L L G I G I

CCAACCAAGGGGTATCCTTACTCCTCACTAGAGTACTTCGAAAAAGCCCTGGAGGAGCTGGCAQGAGGTTCCGGTGTGGCTGGTAGGTTAGGTTAGGCTGGC
200
GGTTGGTCCCCATAGGAAATGAGGAGTGTCTCATGAAGCTTTTGGGACCTCTCGACCGTGTCAAAGGCCACTACCGACCAATCCAAGGCC
hTCC1

P N Q G G I L Y S S L E Y F E K A L E E L A A A A F P G O G W L G S A

CGCGGACAAATACGCCGCAAAAACCGCAACACGTGAATTTTTCAGGAACCTGGAGACCTCGATCGTCAGCTCATCGCTGATCCACGACCGAGGCC
300
CGCCTGTTATGCGGCCGTTGGCGTTGGTCACTTAAAGGTCCTGACCGTCTGGAGCTAGCAGTCGAGTAGTGGACTAGGTGCTGGTCCGG
hTCC1

A D K Y A G K N R N H V N F F Q E L A D L O R Q L I S L I H D Q A

AACGGGGTCCAGACGACCGCGACATCCTGGAGGGCGCAAGAAAGGTCTCGAGTTCTGGCGCCCGTGGCTGTGGACCTGACCTACATCCGGTCGTCG
400
TTGCGCCAGGTCTGCTGGCGCTGTAGGACCTCCCGCGTTCTTCCAGAGCTCAAGCACGGGGCACCGACTGGACTGGATGTAGGGCCAGCAGC
hTCC1

N A V Q T T R D I L E G A K K G L E F V R P V A V O L T Y I P V V

GGCACGGCCATCGGCCGCTTCCAGGCGCCGTTGGCGGGCGCAAGAAAGGTCTCGAGTTCTGGCGCCCGTAAAGCTTGCTACTTGGTCGAAACGCT
500
CCGTGGGGATAGCCGGCGGAAGGTCCCGCGAAAACGCCCGCCGCTACCGGCATCACCCGCCGCGAAATTCAACGGATGAACCGAGCACTTFCGCA
hTCC1 Hind3 DELETED

G H A L S A A F Q A P F C A G A M A V V G G A L K L A Y L V V K T L

GATCAACCGGAAGCTTACTCAACTCCTCAAATTGCTTGCCTAAATTGGCGAGTTGGTCGGCCGCGCATTGGGACATCATTCGGATGTGGCGGACATC
600
CTAGTTCGCTTCGAATGAGTTGAGGAGTTAACGAACGGTTAACCGCTCAACCAGCGCCGGTAACGCCGTAGTAAGCCTACACCGCTGTAG
hTCC1

■ DELETED ■ Hind3 ■ DELETED ■

I N A K L T O L L K L L A K L A E L V A A A I A O I I S O V A O I

ATCAAGGGCATCTCGGAGAAGTGTGGAGTTCATCACAAACCGCCTCAACGGCTGAAAGAGCTTGGACAAGCTCACGGGGGGGTGACCGGACTGT
700
TAGTTCCCGTAGGAGCCTTCACACCCCTCAAGTAGTGTGCGAGTTGGCGCTGGTCTCGAACACCGCTGTAGTAAGCCTACACCGCTGTAG
hTCC1

I K G I L G E V W E F I T N A L N G L K E L W D K L T G W V T G L

TCTCTCGAGGGTGGTCGAACCTGGAGTCCTTCTTGGCGGGCTCCCGGGCTTGACCGGGCGACCGAGGGCTTGTCGCAAGTGACTGGCTTGGCTG
800
AGAGAGCTCCACCAAGCTTGGACCTCAAGGAAGAAACGCCGAGGGGCCACTGGCCGCGTGGTCCGGAAACAGCGTCACTGACCGAACAGCCACG
hTCC1

F S R G W S N L E S F F A G V P G L T G A T S G L S Q V T G L F G A

GGCCGGTCTGTCGCGATCGTCGGCTTGCTCACGGGATAGCCTGGCGAGCTAGCCAGCTTGGCGCCCTGGCCGGCATTGGGGGGGGTCCGGTTTT
900
CGGGCCAGACAGGGTAGCAGCCGAACCGAGTGCCTATCGGACCGCTCGAGTCGGTCAACGGGGGGACCGGGCGTAACCCCGCCAGGCCAAA
hTCC1

A G L S A S S G L A H A D S L A S S A S L P A L A G I G G G S G F

FIG. 9a

Sheet 1 of 2

Thursday, July 22, 1999 1:35 PM

Page 2

TCG1(-TD.1).mod. (1 > 1225) Site and Sequence

CGGGGCTTGCGAGCCCTGGCTCAGGTCCATGCCGCCTCAACTCGGCAGGGCGCTACGGCCCCCGAGCTGATGGCCCCCTCGCGCCCCCTGCCGAGCAGGTGCG

1000

CCCCCGAACGGCTCGGACCGAGTCCAGGTACGGCGAGTTGAGCCGATGCCGGGGCTCGACTACCCGGCGAGCCCGGGACGGCTCGTCCAGC

HTCC1

G G L P S L A Q V H A A S T R C A L R P R A D G P Y G A A A E Q V

CGGGGAGTCGAGCTGGTCTCGCGCAGGGTCCAAAGGTATGGCGGACCCGTAGGCATGGCGGGCATGCCCGCTTCCGGGGCGTCGAAGGGAC

1100

CGCCCGTCAGCGTCCGACCAAGAGCGCGTCCAAAGGGTCCATACCCGGCTGGCATCGTACCCGGCGTACGTGGGAGAAGCCCCCGCAGCTTCCCTG

HTCC1

G G Q S Q L V S A Q G S O G M G G P V G M G G M H P S S G A S K G T

GACGACGAAGAACTACTCGGAAGGCAGGGCGGGCGCCACTGAAGACGCCAGGGCGCCAGTCGAAGCTGACGCCGGCGTGGCAAAGGTGCTGGTA

1200

CTGCTGCTTCTTCATGAGCCTTCCGCGCCGCCGCGGTGACTTCTGCCTGCGCGCGGTCACTTCGACTGCCCGCCACCCGTTTCCACGACCAT

HTCC1

T T K K Y S E G A A A G T E D A E R A P V E A D A G G G Q K V L V

CGAACCGTCGTCTAACGGCGAATT

1225

GCTTTCAGCAGATTGCCGCTTAAG

HTCC1 EcoRI

R N V V . R R I

09586622-302000

FIG. 9a

Sheet 2 of 2

T Cell Epitope Mapping of HTCC-1

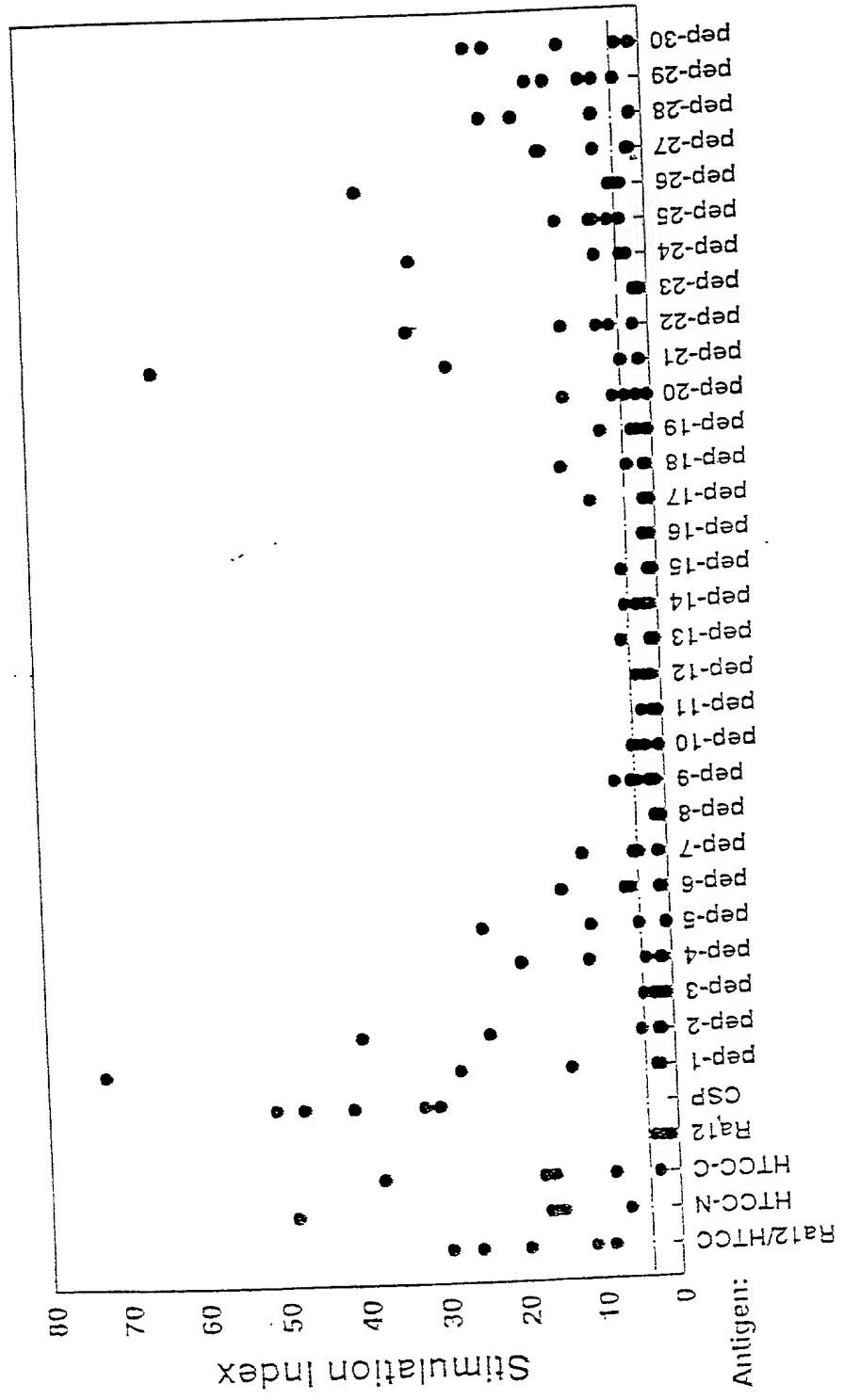


Fig. q

Friday, July 23, 1999 8:41 AM

nTCC1(fl)-TM2 Map.MPO (1 > 1225) Site and Sequence

Enzymes : 1 of 515 enzymes (Filtered)

Settings : Circular, Certain Sites Only, Standard Genetic Code

HICL-1 (TM-2)

Page 1

CATATGCATCACCATCACCATGAGCAGAGCTTCATCATCGATCCAACGATCACTGCCATTGACGGCTTGTACGACCTTCTGGGA
GTATACGTAGTGGTAGTGGTAGTGTACTCGTCTCGCAAGTAGTAGCTAGGTTGCTAGTCACGGTAAGTGGCGAACATGCTGGAAAGACCCCT
hTCC1

91

H M H H H H H M S R A F I I O P T I S A I D G L Y O L L G

TTGGAATACCCACCAGGGGTATCCTTACTCCTCACTAGAGTACTTCGAAAAAGCCCTGGAGGAGCTGGCAGCAGCGTTCCGGTGA
AACCTATGGTTGGTCCCCATAGGAAATGAGGAGTATCTCATGAAGCTTTTGGGACCTCTGACCGTGTGCAAAGGCCACT
hTCC1

182

I G I P N Q G G I L Y S S L E Y F E K A L E E L A A A A F P G Q

TGGCTGGTTAGGTTGGCCGGGACAAATACGCCGGCAAAAACCGCAACCACGTGAATTTCAGGAACCTGGCAGACCTCGATCGTCAG
ACCGACCAATCCAAGCCGGCGCTGTTATGCGCCGTTGGCGTTGGTGCACTTAAAAAAGTCCTTGACCGTCTGGAGCTAGCAGTC
hTCC1

273

G W L G S A A D K Y A G K N R N H V N F F Q E L A D L O R Q

CTCATCAGCCTGATCCACGACCAGGCACCGGGTCCAGACGACCCGGACAAGCTTATCCTGGAGGGCGCAAGAAAGGTCTCGAGTTG
GAGTAGTCGGACTAGGTGCTGGTCCGGTTGCGCCAGGTCTGCTGGCGCTGTTGCAATAGGACCTCCCGCGGTTCTTCCAGAGCTCAAGC
hTCC1 Hind3 DELETED

364

P L I S L I H D Q A N A V Q T T R O K L I L E G A K K G L E F

TGGCCCCGGTGGCTGTGGACCTGACCTACATCCGGTGTGGCCACGGCCCTATCGGGCCCTTCCAGGCGCCGTTTGCAGCGGGCGCAT
ACGCGGGCCACCGACACCTGGACTGGATGTAGGGCCAGCAGCCCGTGGGGATAGCCGGCGAAGGTCCCGCGCAAAACCGCCCCCGCTA
DELETED

455

V R P V A V O L T Y I P V V G H A L S A A F Q A P F C A G A M

GGCCGTAGTGGCCCGCGCTTGCCTACTTGGCTGTGAAACGCTGATCAACGCGACTCAACTCCTCAAATTGCTTGCCTAAATTGGCGAG
CCGGCATCACCCGCCCGCGAACGGATGAACCGACACTTGGACTAGTTGCCTGAGTTGAGGAGTTAACGAACGGTTAACCGCCCTC
DELETED

546

A V V G G A L A Y L V V K T L I N A T Q L L K L L A K L A E

TTGGTCGGCCGCCATTGCGGACATCATTCGGATGTGGGGACATCATCAAGGGATCCTGGAGAAGTGTGGAGTTCATCACAAACG
AACCAGCGCCGGGGTAACGCCGTAGTAAAGCTACACCGCTGTAGTAGTTCCCGTAGGAGCCTTCAACACCCCTCAAGTAGTGTGG
DELETED

637

L V A A A I A D I I S O V A D I I K G I L G E V W E F I T N

CGAAGCTTCTCAACGGCCTGAAAGAGCTTGGGACAAGCTCACGGGGTGGGTGACCGGACTGTTCTCTCGAGGGTGGTCGAACCTGGAGTC
GCTTCGAAGAGTTGCCGGACTTCTCGAAACCCCTGTTGAGTGGCCCAACCGCTGACAAGAGAGCTCCACACCCCTCAAGTAGTGTGG
Hind3 hTCC1

728

A K L L N G L K E L W D K L T G W V T G L F S R G W S N L E S

CTTCTTTCGGGGCGTCCCCGGCTTGACCGGGCGGACCGAGCGGCTTGTGCAAGTGACTGGCTTGTGCGGGGGCTGTCCGCATCG
GAAGAAACGCCCGCAGGGGCCGAAGTGGCCCGCTGGTCGCCGAACAGCGTTCACTGACCGAACAGCCACGCCGGCAGACAGGCGTAGC
hTCC1

819

F F A G V P G L T G A T S G L S Q V T G L F G A A G L S A S

FIG. 4d

Sheet 1 of 2

TCGGGCTGGCTCACCGGATAGCCTGGCAGCTCAGCCAGCTTGCCTGGCCGGCATTTGGGGGGGTCCGGTTTGGGGCTTGC
 + 910
 AGCCCCGAACCGAGTCCGCCATCGGACCGCTCGAGTCGGTCQAACGGGGGGGACGGGGCTAACCCCCGGCCAGGCCAAAACCCCCGAACG
 hTCC1
 S G L A H A D S L A S S A S L P A L A G I G G G S G F G G L
 CGAGCCTGGCTCAGGTCCATGCCGCCCTCAACTCGGCAGGCCTACGGCCCCQAGCTGATGGGCCGGTC3GCGCCGCTGCCGAGCAGGTGG
 1001
 GCTCGGACCGAGTCCAGGTACGGCGGAGTTGAGCCGTCCCGATGCCGGGGCTCGACTACCGGGCCAGCCGCCGACGGCTCGTCCAGCC
 hTCC1
 P S L A Q V H A A S T R Q A L R P R A D G P V G A A A E Q V G
 CGGGCAGTCGAGCTGGTCTCCGCCAGGGTTCCAAGGTATGGGGGACCCGTAGGCATGGGCCATGCACCCCTCTCGGGGGCGTCG
 1092
 GCCCGTCAAGCGTCGACCAGAGGGCGCTCCAAAGGGTTCCATACCCGCCCTGGGCATCCGTACCCGCCGTACGTGGGGAGAAGCCCCCGCAGC
 hTCC1
 G Q S Q L V S A Q G S C G M G G P V G M G G M H P S S G A S
 AAAGGGACGACGACGAAGAAGTACTCGGAAGGCGCGGGCACTGAAGACGCCAGCGCGCCAGTCGAAGCTGACGGCGGGCGTG
 1183
 TTTCCCTGCTGCTGCTTCTTCATGAGCCTCCGCCGCCGTGACTTCTGGGGCTCGCGCGCGTCAAGCTTCGACTGCGCCGCCAC
 hTCC1
 K G T T T K K Y S E G A A A G T E D A E R A P V E A O A G G
 GGCAAAAGGTGCTGGTACGAAACGTCGTCTAACGGCGAATT
 1225
 CCGTTTCCACGACCATGCTTGAGCAGATTGCCGCTTAAG
 hTCC1 EcoRI
 S S K Y L W S N V V S S I

FIG. 4d

Sheet 2 of 2

Monday, July 26, 1999 3:25 PM

ht(184-392)-HS-ht(1-129).mpd (1 > 2232) Site Sequence
Enzymes : 3 of 515 enzymes (Filtered)

Settings: Linear, Certain Sites Only, Standard Genetic Code

HICL=4134-21X-1011-1-1 ~ Page 1

CATATGCATCACCATCACCATCACCATGTCGGGGACATCAAGGGCATCCTCGGAGAAGTGTGGGAGTTCATCACAAACGCCCTCAACGGGCTGAAAG
GTATACGTAGTGGTAGTGGTAGTGCTACACCGCTGTAGTAGTTCCCGTAGGGAGCCTCTTCACACCGCTCAAGTAGTGTGCGCGAGTTGCCGGACTTT
Met / HIS TAG hTCC1 (184-392)
H M H H H H H H O V A D I I K G I L G E V W E F I T N A L N G L K
AGCTTGGGACAAGCTCACGGGTGGTGAACGGACTGTTCTCGAGGGTGGTGAACCTGGAGTCCTTTCGGGGCGTCCCCGGCTTGACCGGGCG
TCGAAACCCGTTCGAGTGCCCCACCCACTGGCTGACAAGAGAGCTCCACCGAGCTGGACCTCAGGAAGAAAC3CCCCGAGGGGGCGAACCTGGCCCGCG
hTCC1 (184-392)
E L W D K L T G W V T G L F S R G W S N L E S F F A G V P G L T G A
GACCAGCGGTTGTCGCAAGTGAUTGGCTTGGTGCACGGCTGTGTCGCATCGTCGGCTTGGCTCACGGGATAGCCTGGCGAGCTAGCCAGC
CTGGTGCACGAGCGTTCACTGACCGAACAGCCACGCCGGCCAGACAGGGTAGCAGGGGGAAACCGAGTGCCTATCGGACCGCTCGAGTCGGTCG
hTCC1 (184-392)
T S G L S Q V T G L F G A A G L S A S S G L A H A O S L A S S A S
TTGCCCGCCCTGGCCGGCATTGGGGCGGGTCCGGTTGGGCTTGCCGAGCCTGGCTCAGGTCATGCCGCTCAACTCGGCAGGGCTACGGCCCC
AACGGGGGGGACGGGGCGTAACCCCCGGCCAGGCCAAAACCCCCGAAACGGCTCGGACCGAGTCCAGGTAACGGCGAGTTGAGCCGTCGCGATGCCGGGG
hTCC1 (184-392)
L P A L A G I G G G S G F G G L P S L A Q V H A A S T R Q A L R P
GAGCTGATGGCCCCGGTCGGCCCGCTGCCGAGCAGGTGGCGGGCAGTCGCAGCTGGCTCCGCCAGGGTTCCAAGGTATGGCGGACCCGTAGGCAT
CTCGACTACCGGGCCAGCCGCCAGGGCTCGCCAGCCCGCTCAGGCGCCCTGCCAGGCTCCATACCCGCTGGCATCCGTA
hTCC1 (184-392)
R A D G P V G A A A E Q V G G Q S Q L V S A Q G S Q G M G P V G M
GGCCGGCATGACCCCCCTTCCGGGGCGTCGAAAGGGACGACGCCAGAAGACTCGGAAGGGCGGGCGGGCACTGAAGACGCCGAGCGCGCGCA
CCCCCGTACGTGGGAGAACCCCCCGCAGCTTCCCTGCTGCTCTTCATGAGCCTTCCCGCCCGCCCGTACTCTCCGGCTCGCGCGCGT
hTCC1 (184-392)
G G M H P S S G A S K G T T T K K Y S E G A A A G T E D A E R A P
GTCGAAGCTGACGCCGGCGGTGGCAAAAGGTGCTGGTACGAAACGTCGTCGAATTCTGGTGGATTTCGGGGCGTTACCGGAGATCAACTCCGCGA
CAGCTCGACTGCCGCCGCCACCGTTTCCACGACCATGCTTGCAGCAGCTTAAGTACCACTAAAGCCCCGCAATGGTGGCTCTAGTTGAGGGCGT
hTCC1 (184-392) EcoRI TbH9
V E A D A G G G Q K V L V R N V V E F M V D F G A L P P E I N S A
GGATGTACGCCGGCCCGGGTTCGGCTCGCTGGTGGCCGGCTCAGATGTGGACAGCGTGGCGAGTGACCTGTTTCGGCCCGTGGCGTTCTG
CCTACATGCCGCCGGCCAAGCCGGAGCACCACCGCCGAGTCTACACCCGTGCGACCGCTACTGGACAAAAGCCGGCGAGCCGCAAAGTCAG
TbH9
R M Y A G P G S A S L V A A A Q M W O S V A S D L F S A A S A F Q S
GGTGGTCTGGGGTCTGACGGTGGGTGCTGGATAGGTTGTCGGCGGGCTGATGGTGGCGGGCTCGCCGTATGTGGCGTGGATGAGCGTCACCGCG
CCACCAAGACCCAGACTGCCACCCAGCACCTATCCAAGCAGCCGCCAGACTACCAACCGCCGCGAGCGGATACACCGCACCTACTCGCAGTGGCGC
TbH9
V V W G L T V G S W I G S S A G L M V A A A S P Y V A W M S V T A

FIG. 10

Sheet 1 of 3

GGGCAGGGCGAGCTGACCCGGCCAGGTCCGGGTTCTCGGCGGCCATCGAGACGGCGTATGGGCTGACGGTGCCCCCGGGTGAACGGAGAACCG
CCCGTCCGGCTCGACTGGGGCGGTCCAGGCCAACGACGCCCGGGATGCTCTGCCGATACGGACTGCCACGGGGCACTAGGGCTCTGG
TbH9
G Q A E L T A A Q V R V A A A A Y E T A Y G L T V P P P V I A E N
GTGCTGAACTGATATTCTGATAAGGACCAACCTCTTGGCAAAACACCCGGCGATCGCGGTCAACGAGGCCAATACGGAGATGTGGGCCAAGA
CACCACTTGACTACTAAGACTATCGCTGGTGGAGAACCCGGTTTGTCGGGGCGTAGCGCCAGTTGCTCGGCTATGCCGCTACACGGGGTCT
TbH9
R A E L M I L I A T N L L G Q N T P A I A V N E A E Y G E M W A O D
CGCCGCCGGATGTTGGCTACGCCGGCGACGGCGACGGCGACGGCTGCGTTGGAGGAGGCCAGATGACCAGCAGGGTGGCTC
GCGCGGGCTACAAACCGATGCCGGCCGCTGCCGCTGCCCTGCCAACGACGGCAAGCTCTCCCGGGCTACTGGTGGGGCACCCGAG
TbH9
A A A A M F G Y A A A T A T A T L L P F E E A P E M T S A G G L
CTGAGGAGGCCGCCGGTCAAGGAGGCCCGACACCCCGGGCGAACAGTTGATGAAACAATGTCGGCCAGGCCAGCTGGCCAGCCA
GAGCTCGTCCGGCGCCAGCTCCGGAGGCTGTGGGGCGCGCTGGTCAACTACTTGTACACGGGGTCCGCGACGTTGTCGACCGGGTGGGT
TbH9
L E Q A A A V E E A S D T A A A A N Q L M N N V P Q A L G Q L A Q P
CGCAGGGCACCGCCTTCTCCAAGCTGGGTGGCTGTGGAAGACGGCTCGCCGATCGGTGCGGATCAGCAACATGGTGTGATGGCAACACCA
GCGTCCCGTGGTGGAGAACAGTTGACCCACCGGACACCTTCTGCCAGAGCGGGTAGCCAGCGGCTAGTCGTGTACACAGCTACCGGTTGGT
TbH9
T O G T T P S S K L G G L W K T V S P H R S P I S N M V S M A N N H
CATGCGATGACCAACTCGGGTGTGCGATGACCAACACCTTGAGCTCGATGTTGAAGGGCTTGCCTCCGGGGCGCCAGGGCGTGCAACACCGG
GTACAGCTACTGGTTGAGCCCACACAGCTACTGGTTGAGACTCGAGCTACAACCTCCGAAACGAGGCCGCCGGGTCGGCACGTTGGCG
TbH9
M S M T N S G V S M T N T L S S M L K G F A P A A A A Q A V Q T A
GCCCAAAACGGGGTCCGGCGATGAGCTCGCTGGCAGCTCGCTGGTTCTCGGGCTGGCGTGGGTGGCCCAACTTGGTGGCGCTCGG
CGCGTTTCCCCAGGCCGCTACTCGAGCGACCCGTCAGCGAACCAAGAACGCCAGACCCGCCACCCACCGGGGTGAACCCAGGCCGGAGCC
TbH9
A Q N G V R A M S S L G S S L G S S G L G G G G V A A N L G R A A S
TCGGTTGTTGTCGGTCCGGCAGGCCCTGGGGCGGCCAACAGGCAGTCACCCCGGGCGCGGGCGCTGCCGCTGACCGCTGACCGCCGG
AGCCAAGCAACAGCCACGGCGTCCGGACCCGGCGCTGGTCCGTCAGTGGGGCGCCGCGCCGGACGGGACTGGTGGACTGGTGGCGCTCGG
TbH9
V G S L S V P Q A W A A A N Q A V T P A A R A L P L T S L T S A A E
AAGAGGGCCCGGGCAGATGCTGGCGGGCTGCCGGTGGGAGATGGCGCCAGGGCCGGTGGTGGCTCAGTGGTGTGCTGCGTGTCCGCCGACCC
TTCTCCCGGGCCGCTACGACCCGCCAACGGCCACCCCGTACCCCGGGTCCGGCACCCGAGTCACACAGCGCACAGGCCGCTGGG
TbH9
R - G P G Q M L G G L P V G Q M G A R A G G G L S G V L R V P P R P
TATGTGATGCCGATTCTCGGCAGCCGGCATATGAGCAGAGCGTTCATCATCGATCCAACGATCGTGCCTGACGGCTGTACGACCTTCTGG
ATACACTACGGCGTAAGAGGCCGTCGGCGCTATAGTACTCGTCTCGCAAGTAGTAGCTAGGTGCTAGTCACGGTAACGCCAACATGCTGGAAAGACC
TbH9
RV
HTCC1 (1-129)
Y V M P H S P A A G O I M S R A F I I O P T I S A I D G L Y D L L

FIG. 10

Sheet 2 of 3

GGATTGGAATACCCAAACCAAGGGGGTATCCTTTACTCCTCACTAGAGTACTCGAAAAAGCCCTGCAGGGAGCTGGCAGCAGCGTTCCGGGTGATGGCTG
CCTAACCTTATGGGTTGGTCCCCCATAGGAAATGAGGAGTGATCTCATGAAGCTTTTCGGGACCTCCTCGACCGTCTGTCQCAAAAGGGCCACTACCGAC
hTCC1 (1-129) 2000

G I G I P N Q G G I L Y S S L E Y F E K A L E E L A A A F P G D G W
GTAGGTTCCGGCGCGGACAAATACGCCGGCAAAAACGCCAACCGTGAATTTTTCCAGGAACCTGGCAGACCTCGATCGTAGCTCATCAGCCTQATC
CAATCCAAGCCGGCGCTGTTATGCCGGTTGGCGACTTAAAAAAGGTCTTGACCGTCTGGAGCTAGCAGTCAGTAGTCGGACTAG
hTCC1 (1-129) 2100

L G S A A D K Y A G K N R N H V N F F D E L A D L D R Q L I S L I
CACGGACCAGGCCAACGGGTCCAGACGACCCGGACATCCTGGAGGGGCCAAGAAAGGTCTCGAGTTCGTGCGCCGGTGGCTGTGGACCTGACCTACA
GTGCTGGTCCGGTTGCCGCCAGGTCTGCTGGCGCTGTAGGACCTCCCGGGTTCTTCCAGAGCTAACGACGCCGGCACCGACACCTGGACTGGATGT
hTCC1 (1-129) 2200

H O Q A N A V Q T T R D I L E G A K K G L E F V R P V A V D L T Y
TCCC GGTCGT CGGGCACGCCCTATAAGATATC 2232
AGGGCCAGGCCCGTGC GGATATTCTATAG
hTCC1 (1-129) RV

I P V V G H A L . O I

bioRxiv preprint doi: <https://doi.org/10.1101/000000>

FIG. 10

Sheet 3 of 3

Monday, July 26, 1999 2:42 PM

Page 1

ht(1-149)-H9-ht(161-392).mod (1 > 2365) Site and Sequence
Enzymes : 3 of 515 enzymes (Filtered)
Settings : Circular, Certain Sites Only, Standard Genetic Code

CATATGCATCACCATCACCATCACATGAGCAGAGCGTTCATCATCGAACGATCAGTGCCATTGACGGCTTGACGACCTCTGGGGAA
GTATACTAGTGGTAGTGGTAGTGTACTCGTCTCGCAAGTAGTAGCTAGGTTGCTAGTCACGGTAACCTGCCAACATCGCTCGAAGAGCCCCCT
Met / HIS TAG ————— nTCC1 (1-149) —————
H M H H H H H M S R A F I I D P T I S A I D G L Y O L L G
TTGGAATACCCAACCAAGGGGGTATCCTTACTCCTCACTAGAGTACTTCGAAAAAGCCCTGGAGGGAGCTGGCAGCAGCGTTCCGGGTGA
AACCTTATGGGTTGGTCCCCCATAGGAAATGAGGAGTGTCTCATGAAGCTTTTCCGGACCTCCTCGACCGTCGCAAAAGGCCACT
nTCC1 (1-149) —————
I G I P N C G G I L Y S S L E Y F E K A L E E L A A A A F P G D
TGGCTGGTTAGGTTGGCCGGACAAATACGCCGGCAAAAACCGCAACCACGTGAATTTCAGGAACCTGGCAGACCTCGATCGTCAG
ACCGACCAATCCAAGCCGGCGCTGTTATGCGGCCGTTGGCGTTGGTGCACTTAAAAAGGTCTTGACCGTCGGAGCTAGCAGTC
nTCC1 (1-149) —————
G W L G S A A D K Y A G K N R N H V N F F Q E L A D L O R Q
CTCATCAGCCTGATCCACGACCAGGCCAACGCCGGTCCAGACGACCGCCGQACATCCTGGAGGGGCCAAGAAAGGTCTCGAGTTCGTGGCGCC
GAGTAGCGGACTAGGTGCTGGTCCGGTTGCGCCAGGTCTGCTGGCGCTGTAGGACCTCCCGCGGTTCTTCAGAGCTCAAGCACCGCG
nTCC1 (1-149) —————
L I S L I H D Q A N A V G T T R O I L E G A K K G L E F V R
CCGTGGCTGTGGACCTGACCTACATCCCGTCGTCGGCACGCCCTATCGGGCCGCTTCAGGCAGGCTTTGGCGGGCGCGATEGGCGT
GCCACCGACACCTGGACTGGATGTAGGGCCAGCACCCGTCGGGGATAGCCGGCGGAAGGTCCCGCGAACAAACGCCCGCGETACCGGCA
nTCC1 (1-149) —————
P V A V D L T Y I P V V G H A L S A A F Q A P F C A G A M A V
AGTGGGCGGCCTTAAGCTTATGGGGATTTCGGGGCGTTACCAACCGGAGATCAACTCCCGGAGGATGTACGCCGGCCGGGTTGGCC
TCACCCGCCCGCGAATTGAAATACCAACCTAAAGCCCCGCAATGGTGGCTCTAGTTGAGGGCGCTCCTACATGCGGCCGGGCCCAGCCGG
hTCC1 (1-149) ————— Hind3 ————— TbH9 —————
V G G A L K L M V D F G A L P P E I N S A R M Y A G P G S A
TCGCTGGTGGCCGGCTCAGATGTGGACAGCGTGGCAGTGACCTGTTTCGGCCGCTCGGGCGTTTCAGTCGGTGGTCTGGGGTCTGA
AGCGACCAACCGCCGCGAGTCTACACCCGTGCGACCGCTCACTGGACAAAGCCGGCGAGCCGCAAAGTCAGCCACCAAGCCCCAGACT
TbH9 —————
S L V A A A Q M W D S V A S D L F S A A S A F Q S V V W G L
CGGTGGGTCGTGGATAGGTTCTCGGGGGCTGATGGTGGCGGTCTCGCCGTATGGCGTGGATGAGCGTCACCGCGGGCAGGC
GCCACCCAGCACCTATCCAAGCACCCGAGACTACCAACCGCCAGAGCGGCATACACCGCACCTACTCGCAGTGGCGCCCGTCCG
TbH9 —————
T V G S W I G S S A G L M V A A V S P Y V A W M S V T A G Q A
CGAGCTGACCGCCGCCAGGTCCGGGTTGCTGCGGCGGCTACGAGACGGGTATGGGCTGACGGTGCCCCCGCCGGTATCGCCGAGAAC
GCTCGACTGGCGGGTCCAGGCCAACGACGCCGGATGCTCTGCCGATAACCGACTGCCACGGGGCGCCACTAGCGGCTCTTG
TbH9 —————
E L T A A Q V R V A A A A Y E T A Y G L T V P P P V I A E N

FIG. 11

Sheet 1 of 3

CGTGCTGAACTGATGATTCTGATAAGCCACCAACCTCTGGGGCAAACACACCCCCGGCGATCGCGGTCAACGAGGCCGAATACGGCGAGATGT
GCACGACTTQACTACTAAGACTATCGCTGGTCCAGAACCCCGTTTGTGGGGCCGCTAGCGCCAGTTGCTCCGGTTATGCCGCTCTACA

TbH9

910

R A E L M I L I A T N L L G Q N T P A I A V N E A E Y G E M

GGGCCCAAGACGCCGCCGCGATGTTGGCTACGGCGCGACGGCGACGGCGACGGCGACGTTGCTGCCGTTCGAGGAGGCCGGAGAT
CCCGGGTTCTGCGGCGGCCGCTACAAACCGATGCCGCCGCCGCTGCCGCTGCCGCTGCCAACGACGGCAAGCTCCTCCGCCGCTCTA

TbH9

1001

W A Q D A A A M F G Y A A A T A T A T A T L L P F E E A P E M
GACCAGCGCGGGTGGGCTCCTCGAGCAGGCCGCCGCCGCTCGAGGAGGGCTCCGACACCGCCGCCGGAACCAAGTTGATGAACAATGTGCC

1092

CTGGTCGGCGCCACCCGAGGAGCTCGTCCGGCGGGCGGCTCCGGAGGGCTGTCGGCGCCGCTGGTCAACTACTTGTACACGGG

TbH9

T S A G G L L E Q A A A V E E A S O T A A A N Q L M N N V P

CAGGGCGCTGCAACAGCTGGCCCAGCCCACGCAGGGCACCCACGCCCTCTCCAAGCTGGGTGGCCTGTGGAAGACGGCTCGCCGATCGGT
GTCCCGCACGTTGTCGACCGGGTCGGGTGCGTCCCGTGGTGCAGGAGAGGTTGACCCACCGGACACCTTCTGCCAGAGCGGGTAGCCA

1183

TbH9

Q A L Q Q L A C P T Q G T T P S S K L G G L W K T V S P H R

CGCCGATCAGCAACATGGTGTGATGGCCAACAACACATGTCGATGACCAACTCGGGTGTGATGACCAACACCTTGAGCTCGATGTT
GCGGCTAGTCGTTGTACACAGCTACCGGTTGTTGGTACAGCTACTGGTTGAGCCCACACAGCTACTGGTTGGAACCTCGAGCTACAA

1274

TbH9

S P I S N M V S M A N N H M S M T N S G V S M T N T L S S M L

GAAGGGCTTTGCTCCGGCGGCCGCCAGGGCGTGCACACCGCCGGCGAAACGGGGTCCGGCGATGAGCTCGCTGGCAGCTCGTG
CTTCCCGAAACGAGGCCGCCGCCGGCGGGTCCGGCACGTTGGCGCCGTTTGCCCCAGGCCGCTACTCGAGCGACCCGTCAGCGAC

1365

TbH9

K G F A P A A A A Q A V Q T A A Q N G V R A M S S L G S S L

GGTTCTCGGGCTGGGGCGTGGGGTGGCCCAACTGGGTGGGGCGGCTCGTGGTGTGGTGGTGGCTGGCCAGGGCTGGCCCGCG
CCAAGAAGCCCAGACCCGCCACCCGCCAGGGCGGGTTGAACCCAGGCCGCCGGAGCCAGCCAAGCAACAGCCACGGCGTCCGGACCCGGCGCC

1456

TbH9

G S S G L G G G V A A N L G R A A S V G S L S V P Q A W A A

CCAACCAGGCAGTCACCCCGGCCGCCGCTGCCGCTGACCAAGCCTGACCCAGCGCCGCCGAAAGAGGGCCCGGGAGATGCTGGCGG
GGTTGGTCCGTCAGTGGGGCGCCGCCGCCGGAGCGGGACTGGTCCGACTGGTCCGCGCTTCTCCCGGGCCGCTACGACCCGCC

1547

TbH9

A N Q A V T P A A R A L P L T S L T S A A E R G P G Q M L G G

GCTGCCGGTGGGGCAGATGGGGCCACGGGCCGGTGGTGGGCTCAGTGGTGTGCTGCCGCCGACCCCTATGTGATGCCGCATTCT
CGACGGCCACCCCGTCTACCCGCCGGTCCCGCCACCCGAGTCACCCACACGACGACACAAGGCCGCTGGGATACACTACGGCGTAAGA

1638

TbH9

L P V G Q M G A R A G G G L S G V L R V P P R P Y V M P H S

FIG. 11

Sheet 2 of 3

CCGGCAGCCGGCAAGCTTACTCAACTCCTCAAATTGCTTGCCAAATTGGCGGAGTTGGTCGCGGCCATTGGACATCATTCGGATG
 1729
 GGCGTGGCCGTTGAATGAGTTGAGGAGTTAACGAAACGGTTAACCGGCTCAACCAGEGCCGGCGTAACGGCTGTAGTAAAGGCTAC
 — TaHe — Hind3 — hTCC1 (161-392) —
 P A A G K L T Q L L K L L A K L A E L V A A A I A G I I S D
 TGGCGGACATCATCAAGGGCATCCTCGGAGAAGCTGCGGAGTTCATCACAAACCGCTCAACGGCTGAAAGAGCTTTGGGACAAGGTCAC
 ACCGCCTGTAGTAGTTCCCGTAGGAGCCTCTCACACCCCTCAAGTAGTGTGTTGCCGAGTTGCCGACTTTCTCGAAACCCGTGTTCCGAGTG
 1820
 — hTCC1 (161-392) —
 V A D I I K G I L G E V W E F I T N A L N G L K E L W O K L T
 GGGGTGGGTGACCGGACTGTTCTCTCGAGGGTGGTCGAACCTGGAGTCCTTCTTGGGGCTCCCCGGCTTGACCGGGCGACCAAGGGC
 1911
 CCCCCACCCACTGGCCTGACAAGAGAGCTCCACCAAGCTTGGACCTCAGGAAGAAACGCCCGCAGGGGGCAACTGGCCCGCTGGTCGGCG
 — hTCC1 (161-392) —
 G W V T G L F S R G W S N L E S F F A G V P G L T G A T S G
 TTGTGCGCAAGTGACTGGCTTGGTGGTGGCGCTGTCGCGCATCGTCGGCTGGCTCACCGGGATAGCCTGGCGAGCTCAGCCAGCT
 2002
 AACAGCGTTCACTGACCGAACAGCCACGCCGGCAGACAGGCAGCAGCCCQAACCGAGTGCCTATGGACCGCTCGAGTCGGTCGA
 — hTCC1 (161-392) —
 L S Q V T G L F G A A G L S A S S G L A H A D S L A S S A S
 TGCCCGCCCTGGCCGGCATTGGGGCGGCTCGGTTTGGGGCTTGGCGAGGCTGGCTCAGGTCCATGCCCTCAACTCGGCAGGGCT
 2093
 ACGGGCGGGACCAGGGCTAACCCCGCCAGGCCAAAACCCCGAACGGCTGGACCGAGTCCAGGTACGGGGAGTTGAGCCGTCCCGA
 — hTCC1 (161-392) —
 L P A L A G I G G G S G F G G L P S L A Q V H A A S T R Q A L
 ACGGGCCCCGAGCTGATGGCCCGCTGGCGCCGCTGCCGAGCAGGTGGCGGGCAGTCGCAGCTGGCTCCGGCGAGGGTTCCCAAGGTATG
 2184
 TGCCGGGGCTGACTACCGGGCCAGCCGGACGGCTCGTCCAGCCGGCTCAGCGTCCACCAAGAGGGCGTCCCAAGGGTTCCATAC
 — hTCC1 (161-392) —
 R P R A D G P V G A A A E Q V G G Q S C L V S A Q G S Q G M
 GCGGACCCGTAGGCATGGCCGGCATGCACCCCTTCTGGGGCGTGCAGAAAGGGACGACGCCAAGAAGTACTCGGAAGGGCGGGCGG
 2275
 CGCGCTGGGCTCCGTACCCGCCGTACGTGGGAGAACGCCCGCAGCTTCCCTGCTGCTGCTTCAATGAGCCTTCCCGCGCCGCC
 — hTCC1 (161-392) —
 G G P V G M G G M H P S S G A S K G T T T K K Y S E G A A A
 GCACTGAAGACGCCGAGCGCGCCAGTCGAAGCTGACGCCGGCGGTGGCAAAAGGTGCTGGTACGAAACGTCGTCTAACGGCGAATT
 2365
 CGTGACTCTGGGGCTCGCCGCCGGTCAAGCTCCACTGCCGCCACCCGTTCCACGACCATGCTTGCAGCAGATTGGCGCTTAAG
 — hTCC1 (161-392) — EcoRI —
 G T E O A E R A P V E A O A G G G Q K V L V R N V V . R R I

FIG. 11

Sheet 3 of 3

Monday, July 25, 1989 2:45 PM

ht(184-392)-h9-h9(1-200).mpd (1 > 2446) Site Sequence

Enzymes : 3 of 315 enzymes (Filtered)

Settings: Linear, Certain Sites Only, Standard Genetic Code

Page 1

CATATGGCATCACCATCACCATCACCATGCTGGGACATCATCAAGGGCATCCTCGGAGAAAGTGTGGAGATTCATCACAAACGCCQCTAACGGCCTGAAAG

100

GTATACTAGTAGTGGTAGTGGTAGTGTACACCCGCTGTAGTAGTTCCCGTAGGGAGCCTCTTCACACCCCTCAAGTAGTGTGTTGGCGAGTTGCCGGACTTTC

Met / HIS TAG hTCC1 (184-392)

H M H H H H H O V A Q I K G I L G E V W E F I T N A L N G L K

AGCTTGGACAAGCTCACGGGTGGGTGACCGGACTGTTCTCGAGGGTGGTCGAACCTGGAGTCCTCTTGCAGGGCTCCCCGGCTTGACCGGGCG

200

TCGAAACCCGTGTCGAGTCCCCACCCACTGGCCTGACAAGAGAGCTCCACCCAGCTGGACCTCAGGAAGAACGCCCGAGGGCCGAACTGGCCCG

hTCC1 (184-392)

E L W O K U T G W V T G L F S R G W S N L E S F F A G V P G L T G A

300

GACCAGGGCTTGTGCGAACGTGACTGGCTTTCGGTACGGCCGGTCTGTCGCATCGTCGGCTTGGCTCACGGGATAGCCTGGCAGCTCAGCCAGC

CTGGTGGCCGAACAGCGTTCACTGACCGAACAGCCACGGGGGAGACAGGGTACAGCCCCGAAACGGAGTGCACCTATCGGACCGCTGAGTCGGTCA

hTCC1 (184-392)

T S G L S Q V T G L F G A A G L S A S S G L A H A D S L A S S A S

TTGGCCGCCCTGGCCGGCATGGGGGGGGTCCGGTTTGGGGCTTCCGAGCCTGGCTCAGGTCAATGCCGCTCAACTCGGAGGGCTACGGGGGG

400

AACGGGGGGGACGGGGCGTAACCCCCGGCCAGGCCAAAACCCCCGQAACGGCTCGGACCGAGTCCAGGTACGGGGAGTTGAGCCGTCGGCATGGGGGG

hTCC1 (184-392)

L P A L A G I G G G S G F G G L P S L A Q V H A A S T R Q A L R P

500

GAGCTGATGGCCCGGTGGCGCCGCTGGCAGCAGGTGGCGGGCAGTCGCAGCTGGCTCGCGCAGGGTTCCAAGGTATGGGGGACCCGTAGGCAT

CTCGACTACCGGGCCAGCCGGCGACGGCTCGTCAGCCGGCGTCAGCGTCGACCGAGGGCGCTCCAGGGTTCCATACCCGCTGGCATCGCTA

hTCC1 (184-392)

R A D G P V G A A A E Q V G G Q S Q L V S A Q G S Q G M G G P V G M

600

GGGCCGCATGCACCCCTTGGGGGGCTGCAAAGGACGACGAAGAAGTACTCGGAAGGGCCGGGGGGACTGAAGACGCCGAGCCGGGCCA

CCCGCCGTA CGTGGGAGAAGGCCCGCAGCTTCCCTGCTGCTTCATGAGCCTTCCGCGCCGCCGCGTGA CTTCTGGGCTCGGCCGGT

hTCC1 (184-392)

G G M H P S S G A S K G T T T K K Y S E G A A A G T E D A E R A P

GTCGAAGCTGACGGGGGGGGTGGCAAAGGTGCTGGTACGAAACGTCGTCAATTCTGGTGGATTGGGGCGTTACCCACGGAGATCAACTCCGCGA

700

CAGCTTCGACTCGGCCGCCACCCGTTTCCACGACCATGTTCCAGCAGCTTAAGTACCAACCTAAAGGCCCGCAATGGTGGCCCTAGTTGAGGGCGT

hTCC1 (184-392)

EcoRI

TbH9

V E A D A G G Q K V L V R N V V E F M V D F G A L P P E I N S A

GGATGTACGCCGGGGGGTCGGCTGGTGGCGGGCTCAGATGTGGACAGCGTGGCGAGTGACCTGTTTCGGCCGCGTGGCGTTCACTG

800

CCTACATGCCGGGGGGCCAGGCCGGAGCGACCGGGCGAGTCTACACCCGTGCGACCGCTACTGGACAAAGCCGGCGAGCCGCAAAGTCAG

TbH9

R M Y A G P G S A S L V A A A Q M W D S V A S O L F S A A S A F Q S

900

GGTGGCTGGGGCTGACGGTGGGTGCTGAGTTCTGGCGGGCTGATGGTGGCGGGCCCTGCCGTATGTGGCGTGGATGAGCGTCACCGCG

CCACCAAGACCCAGACTGCCACCCAGCACCTATCCAAGCAGCCGCCAGACTACCAACCGCCGCCAGCGCATACACCGCACCTACTCGCAGTGGCGC

TbH9

V V W G L T V G S W I G S S A G L M V A A A S P Y V A W M S V T A

FIG. 12

Sheet 1 of 3

Monday, July 26, 1999 2:45 PM

(184392)-H9-ht(1-200).mod (1 > 2445) Site and Sequence

184392

GGGCAGGCCGAGCTGACCGCCGCCAGGFCGGGTTGCTGGCGGCCCTACGAGACGGCGTATGGCTGACGGTCCCCCGCCGGTGTGCGCGAGAACCGCGTCCGACTGGCGGGGGTCCAGGCCAACGACGCCGGGGATGCTCTGCCCATACCCGACTGCCACGGGGCGGCTACTAGCGGCTCTTGG

1000

TbH9

G Q A E L T A A Q V R V A A A A Y E T A Y G L T V P P P V I A E N

GTGCTGAACGTGATTCTGATAGCGACCAACCTCTGGGCAAAACACCCCGCGATCGCGCTAACGAGGCCAATACGGCAGAGATGTGGGCCAAGCACGACTAAGACTATCGCTGGTTGGAGAACCCGTTTGTGGGGCCGCTAGCGCCAGTTGCTCCGGCTTAFCGGCTCTACACCCGGGTTCT

1100

TbH9

R A E L M I L I A T N L L G O N T P A I A V N E A E Y G E M W A O D

CGCCGCCGCATGTTGGCTACGCCGCCGCCACGGCGACGGCGACGTTGCTGCCQTTGAGGAGGCCGGAGATGACCACGCCGGGTGGGCTCGGGCGCTACAAACCGATGCGGCCGCTGCCGTGCCGTGCAACGACGGCAAGCTCTCCGGCCCTACTGGTGGCGCCACCGAG

1200

TbH9

A A A M F G Y A A A T A T A T A T L L P F E E A P E M T S A G G L

CTCGAGCAGGCCGCCGGTCGAGGAGGCCCTCGACACCCGCCGGCGAACAGTTGATGAACAATGTGCCCCAGGGCTGCAACAGCTGGCCAGGCCAAGCTCGTCCGGGCGCCAGCTCCCTCCGGAGGCTGTGGCCGGCCGTTGGTCAACTACTTGTACACGGGTCCGCACGTTGTCGACCCGGTGGGT

1300

TbH9

L E Q A A A V E E A S D T A A A N Q L M N N V P Q A L Q Q L A Q P

CGCAGGGCACCCACCCCTCTCCAAGCTGGGTGGCTGTGGAAAGACGGCTCGCCGATCGGTGCCGATCAGCAACATGGTGTGATGGCAACACAGCTGGCCAGGCCAAGCTGGTGGGTGGCTCCGGGCGCCGGCCAGGGCTGCAAACACGGCG

1400

TbH9

T Q G T T P S S K L G G L W K T V S P H R S P I S N M V S M A N N H

CATGTCGATGACCAACTCGGGTGTGATGACCAACACCTTGAGCTCGATGTTGAAGGGCTTGCTCCGGGCGCCGGCCAGGGCTGCAAACACGGCG

1500

GTACAGCTACTGGTGAGCCCACACAGCTACTGGTGAGACTCGAGCTACAACCTCCGAAACGAGGCCGGCCGGGGTCCGGCACGTTGGCGC

TbH9

M S M T N S G V S M T N T L S S M L K G F A P A A A A Q A V Q T A

CGCAGGGTCCGGCGATGAGCTCGCTGGCGAGCTCGCTGGTTCTCGGGCTGGCGGGTGGCCGCCACTTGGGTGGCCGGCCAGGGCTGCAAACACGGCG

1600

TbH9

A Q N G V R A M S S L G S S L G S S G L G G G G V A A N L G R A A S

TCGGTTGTTGCGGTGCCGCAGGCCCTGGCCGGCCAACCAAGGCAGTCACCCCGGCCGCAGCCGCTGCGCTGACCGCTGACCGAGCGCCGGAGAAGCC

1700

AGCCAAGCACAGCCACGGCGTCCGGACCCGGCGCCGGTGGTCCGTCACTGGGGCCGCCGGCGGGGGGCGACGGGACTGGTGGACTGGTGGCGCCCT

TbH9

V G S L S V P Q A W A A A N Q A V T P A A R A L P L T S L T S A A E

AAGAGGCCCGGGCAGATGCTGGCGGGCTGCCGGTGGGGCAGATGGGCAGGGGGGGTGGTGGGCTCACTGGTGTGCTGGTGGTCCGGCGGAGCC

1800

TTCTCCGGGGCCGCTACGACCCGCCGACGGCCACCCGGTCACTCCGGGCTCCGGGACCCACCCGAGTCACCAACACGACGCACAAGGGCGCGCTGGG

TbH9

R G P G Q M L G G L P V G Q M G A R A G G G G L S G V L R V P P R P

TATGTGATGCCGCATTCTCGGCAGGCCGGATATCATGAGCAGAGCGTCATCATCGATCCAAAGCATCAGTGGCATTGACGGCTTGACGACCTCTGG

1900

ATACACTACGGCGTAAGAGGCCGTCGGCCGCTATAGTACTCGCTCGCAAGTAGTAGTAGGTTGCTAGTCACGGTAACCTGGCACATGCTGGAAAGACC

TbH9

RV

HTCC1 (1-200)

Y V M P H S P A A G Q I M S R A F I I D P T I S A I D G L Y O L L

FIG. 12

Sheet 2 of 3

Monday, July 26, 1999 2:45 PM

t(134-392)-H9-hf(1-200).mod (1 > 3445) Site 3 Sequence

GGATTGGAATACCCAAACCAAGGGGTATCCTTACTCCTCACTAGAGTACTTCGAAAAAGGCCCTGGAGGGAGCTGGCAGCGAGCTTCCGGGTGATGGCTG 2000

CCTAACCTTATGGGTTGGTCCCCATAGGAAATGAGGAGTGTCTCATGAAQCTTTTCCGGGACCTCCTCGAACCGTCGTGCACAGGGCCACTACCGAC
htCC1 (1-200)

G I G I P N Q G G I L Y S S L E Y F E K A L E E L A A A F P G O G W

GTTAGGTTGGCCGGGACAAAATACGCCGGCAAAACCGCAACCACGTGAATTTTCCAGGAACCTGGCAGACCTCGATCGTCAGCTCATCAGCTGATC 2100

CAATCCAAGCCGGCCCTGTTATGCCGCCGTTGGCTGCACCTTAAAGTCCTTGACCGCTGGAGCTAGCAGTCGAGTAGTCGGACTAG
htCC1 (1-200)

L G S A A D K Y A G K N R N H V N F F D E L A O L D R Q L I S L I

CACGACCAAGGCCAACGGGTCCAGACGACCCGCGACATCGAGGGCGCAAGAAAGGTCTCGAGTTCGTGCGCCCGTGGCTGGACCTGACCTACA 2200

GTGCTGGTCCGGTTGCCAGGTCTGCTGGCGCTGTAGGACCTCCCGGGTTTCCAGAGCTCAAGCACGCCGGCACCGACACCTGGACTQQATGT
htCC1 (1-200)

H O Q A N A V Q T T R O I L E G A K K G L E F V R P V A V O L T Y

TCCCGGTGTCGGCACGCCCTATCGGCCGCTTCCAGGCCGTTTGCAGGGCGATGGCCGTAGTGGCGGCCGCTGGCTACTTGGCTGTGAA 2300

AGGGCCAGCACCCGTGCGGGATAGCCGGCGGAAGGTCCCGGCAAAACGCCGGCGCTACCGGCATCACCCGCCGCGAACGGATGAACCACCACT
htCC1 (1-200)

I P V V G H A L S A A F Q A P F C A G A M A V V G G A L A Y L V V K

AACGCTGATCAACGGACTCAACTCCTCAAATTGCTTGCCTTGGCGAGTTGGTCGCCGCCATTGGGACATCATTCGGATGTGGCGGACATC 2400

TTGCGACTAGTTGGCTGAGTTGAGGTTAACGAGGTTAACCGCTCAACCAGCGCCGGCGTAACGCCGTAGTAAAGCTACACCGCCCTGTAG
htCC1 (1-200)

T L I N A T Q L L K L L A K L A E L V A A A I A O I I S D V A O I

ATCAAGGGCATCCTCGGAGAAGTGTGGAGTTCTAAGATATC
2445

TTGCGACTAGTTGGCTGAGTTGAGGTTAACGAGGTTAACCGCTCAACCAGCGCCGGCGTAACGCCGTAGTAAAGCTACACCGCCCTGTAG

htCC1 (1-200) RV

I K G I L G E V W E F I . O I

FIG. 12

Sheet 3 of 3

Figure : Nucleotide sequence of MTb59

cacgactgcccgaactgaaccgaaacttagtcagcacaaaaccgaagttaggaagagacgaaaagctatggc
 tgagttgacaatccccgtgtatgcacatcccgagcgcaatogaagagtagtacgttaagcttttccacccgc
 cgacaccaggtagagagagaagtcggtaccgtcgatgcggggggacggcatcgacacatcgaggg
 ttgtccatcggtatgcacccaagagactgtcgatgcattcccgggcgaaatctcggtcgcccccaa
 cctcgacgagcacagcgtcggcgccgtgatcctcggtgacttcgagaacatcgaaagaaggtagca
 ggtcaagcgccaccggcgaagtcgttatcggttccgggtcggtcgacgggtttttggggcggttggttaa
 cccgctcgccagccgtcgacggcgccggagacgtcgactccgataactcggtcgccgtggagct
 ccaggcgccctcggtggtgacccggcaaggcgtaaggagccgttgcagacccggatcaaggcgat
 tgcacgcgtatgcacccgtatcgcccgccagcgccagctgtatcatcggtcgaccgcaagacccggcaa
 aaccggccgtcggtcgacaccatcctcaaccagcgccagaactgggagtccggatccaaagaa
 gcagggtgcgtgttatcgtggccatcggtcagaaggaaactaccatcgcccggtacgcccgcac
 acttggaaagaggggcggtgcgtatggactacaccatcgtcgccccggcggtcggtccggccgg
 ttcaaatggcttgcgcgtatcaccgggtcggtcgatcgcccgactgtatgtacgaggcgaaagca
 tgtgtatcatcttcgacgacccgtatcggtcgactaaggcccgaggcataccggcgatctcgatgtgt
 gcgccgtccgccccggcggtggggccatccccggcgatgttgtatctgcattcgccgggttttgg
 gcgctgcgcggaaactgtccgacgtatcggtggcggtcgatcgatccgggtctgcccgtatcgagac
 caaggccaaacgacatctcggtcgatccggatcgatccggatcgatccggatcgatccggatcg
 cctggaaaccgacccgttcaaccaggcggtccggccatcaacgtcggtgtcggtgtcccg
 agtcggcgccggcgcccgatcaaggctatgaaagaggcgccggaaagccctcggttggacccgtt
 gcaataccggcgagctagaagcttcgcgtttcgatcggtcgacgatcgatccggatcgatccgg
 gtggagcgccggcgcccggtggatcgatcgatcgatcgatcgatcgatccggatcgatccgg
 ggagcaagtggtttcgatcttcgtggccaccggcggtcacctggactcggtcgccgtcgaggacgt
 ccggcggttcgaaaccgaattactggaccacatcgccggccatcgatccggatcgatccggatcg
 ccggacagccaaagctcaccggaggaggcccgacaaagctcaccggatcgatccggatcgatccgg
 gaaggggctcgccggccaccgggtggcggtcgatcgatccggatcgatccggatcgatccgg
 ggataagctcgccaaaggaaaggccgtgaaggtcaaaaaggccggcgccgaagaagaagaaatag
 catggatcgccacacttcgcgaactacgcggggcgatccgatcggtcgatccggatcgatccgg
 caaggcccaggagctgtatgcgacatcgccatcgccaggccgatcgatccggatcgatccgg
 gccctacgttttggatcaccggatcgatccaccctggccgtcgatcgatccggatcgatccgg
 gttgt

DRAFT

Figure . : Amino acid sequence of MTb59

MAELTIPADDIQSAIEEYVSSFTADTSREEVGTVDAGDGIAHVEGLPSVMTQELLEFPGGILGVA
LNLDEHSGAVILGDFENIEEGQQVKRTGEVLSVPVGDGFLGRVNPQPLGQPIDGRGDVSDTRRAL
ELQAPSVVHRQGVKEPLQTGIKAIDAMPTIGRGQRQLIIGDRKTGKTAVCVDTILNQRQNWESEGP
KKQVRCVYVAIGQKGTTIAAVRRTLEEGGAMDYTTIVAAAASESAGFKWLAPYTGSAIAQHWMYEG
KHVLIIFFDDLTQQAAYRAISLLRRPPGREAYPGDVFYLHSRLLERCAKLSSDLGGGSLTGLPII
ETKANDISAYIPTNVISITDGQCFLTDLFNQGVRPAINVGVSVSRVCGAAQIKAMKEVAGSLRLD
LSQYRELEAFAAFASDLDAAASKAQLERGARLVELLKQPQSQPMPVVEEQVVSIFLGTGGHLDSPVVE
DVRRFETELLDHMRASEEEILTEIRDSQKLTEEAADKLTEVIKNFKKGFAATGGGSVVPDEHVEAL
DEDKLAKEAVKVKKPAPKKKK

Figure : Nucleotide sequence of MTb82

Figure 1: Amino acid sequence of MTb82

MAKASETERSGPGTQPADAQTATSATVRPLSTQAVFRPDFGDEDNFPHPTLGPDTEPQDRMATTSR
VRPPVRRLLGGGLVEIPRAPDIDPLEALMTNPVVPESKRCPWCNCGRPVGRSDSETKGASEGWCPYCG
SPYSFLPQLNPGDIVAGQYEVKGCIAHGLGWIALALDRNVNGRPVVLKGLVHSGDAEAQAMAMAE
RQFLAEVVHPSIVQIFNFVEHTDRHDPVGYIVMEYVGQSLKRSKGQKL.PVAEAIAYLLEILPAL
SYLHSIGLVYNDLKPEINIMLTEEQLKLIDLGAVERSINSFGYLYGTPGFQAPEIVRTGPTVATDIYT
VGRTLAALTLDLPTRNGRYVDGLPEDDPVLKTYDSYGRLLRAIDPDPRQRFTTAEEMSAQLTGVL
REVVAQDTGVPRPGLSTIFSPSRSTFGV DLLVAHTDVYLDGQVHAEKLTANEIVTALSVPVLDPTD
VAASV LQATVLSQPVQTLDLRAARHGAL DADGVDFSESVELPLMEVRALLGDVAKATRKLDL
AERVGWRWRLVWYRAVAELLTGDYDSATKHFTEVLDTFPGELAPKLALAATAELAGNTDEHKFYQT
VWSTNDGVISAAGLARARSAEGDRVGA VRTLDEVPPTSRHFTTARLTSAVTLLSGRSTSEVTEEQ
IRDAARRVEALPPTEPRVLQIRALVLGGALDWLKDNKASTNHILGF PFTSHGLRLGVEASLRSLAR
VAPTORHRYTLVDMANKVRPTSTF.

09666672-301000

17

Figure : Amino Acid Sequence of secreted DPPD

DPPDPHQPDMTKGYCPGGRWGFGLAVCDGEKYPDGSFWHQWMQTWFTGPQFYFDCVSGGEPLP
GPPPPPGCGGAIPSEQPNAP

DECLARATION

As a below named inventor, I declare that:

My residence, post office address and citizenship are as stated below next to my name; I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural inventors are named below) of the subject matter which is claimed and for which a patent is sought on the invention entitled: **FUSION PROTEINS OF MYCOBACTERIUM TUBERCULOSIS** the specification of which X is attached hereto or _____ was filed on _____ as Application No. _____ and was amended on _____ (if applicable).

I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above. I acknowledge the duty to disclose information which is material to patentability as defined in Title 37, Code of Federal Regulations, Section 1.56. I claim foreign priority benefits under Title 35, United States Code, Section 119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed.

I hereby claim the benefit under Title 35, United States Code § 119(e) of any United States provisional application(s) listed below:

Application No.	Filing Date
60/158,338	October 7, 1999
60/158,425	October 7, 1999

I claim the benefit under Title 35, United States Code, Section 120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, Section 112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, Section 1.56 which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

Application No.	Date of Filing	Status
09/056,556	April 7, 1998	Pending
09/223,040	December 30, 1998	Pending
09/287,849	April 7, 1999	Pending
PCT App. No. WO99/51748	April 7, 1999	

Full Name of Inventor 1:	Last Name: SKEIKY	First Name: YASIR	Middle Name or Initial:
Residence & Citizenship:	City: Seattle	State/Foreign Country: Washington	Country of Citizenship: Canada and Lebanon
Post Office Address:	Post Office Address: 8327 25th Avenue, NW	City: Seattle	State/Country: Washington Postal Code: 98107

Full Name of Inventor 2:	Last Name: REED	First Name: STEVEN	Middle Name or Initial:
Residence & Citizenship:	City: Bellevue	State/Foreign Country: Washington	Country of Citizenship: United States
Post Office Address:	Post Office Address: 2843 122nd Place, NE	City: Bellevue	State/Country: Washington
			Postal Code: 98005
Full Name of Inventor 3:	Last Name: HOUGHTON	First Name: RAYMOND	Middle Name or Initial: L.
Residence & Citizenship:	City: Bothell	State/Foreign Country: Washington	Country of Citizenship: United States
Post Office Address:	Post Office Address: 2636 242nd Place, SE	City: Bothell	State/Country: Washington
			Postal Code: 98021
Full Name of Inventor 4:	Last Name: MCNEILL	First Name: PATRICIA	Middle Name or Initial: D.
Residence & Citizenship:	City: Des Moines	State/Foreign Country: Washington	Country of Citizenship: United States
Post Office Address:	Post Office Address: 1421 S. 248th Street	City: Des Moines	State/Country: Washington
			Postal Code: 98198
Full Name of Inventor 5:	Last Name: DILLON	First Name: DAVIN	Middle Name or Initial: C.
Residence & Citizenship:	City: Issaquah	State/Foreign Country: Washington	Country of Citizenship: United States
Post Office Address:	Post Office Address: 18112 NW Montreux Drive	City: Issaquah	State/Country: Washington
			Postal Code: 98027
Full Name of Inventor 6:	Last Name: LODES	First Name: MICHAEL	Middle Name or Initial: J.
Residence & Citizenship:	City: Seattle	State/Foreign Country: Washington	Country of Citizenship: United States
Post Office Address:	Post Office Address: 9223 36th Avenue, SW	City: Seattle	State/Country: Washington
			Postal Code: 98126

I further declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Signature of Inventor 1 Yasir Skeiky	Signature of Inventor 2 Steven Reed	Signature of Inventor 3 Raymond L. Houghton
Date	Date	Date
Signature of Inventor 4 Patricia D. McNeill	Signature of Inventor 5 Davin C. Dillon	Signature of Inventor 6 Michael J. Lodes
Date	Date	Date

SEQUENCE LISTING

Mtb41 (MTCC#2)

(2) INFORMATION FOR SEQ ID NO:140:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1441 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:140:

GAGGTTGCTG	GCAATGGATT	TCGGGCTTTT	ACCTCCGGAA	GTGAATTCAA	GCCGAATGTA	60
TTCCGGTCCG	GGGCCGGAGT	CGATGCTAGC	CGCCGCGGCC	GCCTGGGACG	GTGTGGCCGC	120
GGAGTTGACT	TCCGCCGCGG	TCTCGTATGG	ATCGGTGGTG	TCGACGCTGA	TCGTTGAGCC	180
GTGGATGGGG	CCGGCGGCCG	CCGCGATGGC	GGCCGCGGCCA	ACGCCGTATG	TGGGGTGGCT	240
GGCCGCCACG	GGCCGCGCTGG	CGAAGGAGAC	GGCCACACAG	GCGAGGGCAG	GGCGGAAAGC	300
GTGGATGGGG	CCGGCGGCCG	CCGCGATGGC	GGCCGCGGCCA	ACGCCGTATG	TGGGGTGGCT	360
GGCCGCCACG	GGCCGCGCTGG	CGAAGGAGAC	GGCCACACAG	GCGAGGGCAG	GGCGGAAAGC	420
GTGGATGGGG	CCGGCGGCCG	CCGCGATGGC	GGCCGCGGCCA	ACGCCGTATG	TGGGGTGGCT	480
GGCCGCCACG	GGCCGCGCTGG	CGAAGGAGAC	GGCCACACAG	GCGAGGGCAG	GGCGGAAAGC	540
GTGGATGGGG	CCGGCGGCCG	CCGCGATGGC	GGCCGCGGCCA	ACGCCGTATG	TGGGGTGGCT	600
GGCCGCCACG	GGCCGCGCTGG	CGAAGGAGAC	GGCCACACAG	GCGAGGGCAG	GGCGGAAAGC	660
GTGGATGGGG	CCGGCGGCCG	CCGCGATGGC	GGCCGCGGCCA	ACGCCGTATG	TGGGGTGGCT	720
GGCCGCCACG	GGCCGCGCTGG	CGAAGGAGAC	GGCCACACAG	GCGAGGGCAG	GGCGGAAAGC	780
GTGGATGGGG	CCGGCGGCCG	CCGCGATGGC	GGCCGCGGCCA	ACGCCGTATG	TGGGGTGGCT	840
GGCCGCCACG	GGCCGCGCTGG	CGAAGGAGAC	GGCCACACAG	GCGAGGGCAG	GGCGGAAAGC	900
GTGGATGGGG	CCGGCGGCCG	CCGCGATGGC	GGCCGCGGCCA	ACGCCGTATG	TGGGGTGGCT	960
GGCCGCCACG	GGCCGCGCTGG	CGAAGGAGAC	GGCCACACAG	GCGAGGGCAG	GGCGGAAAGC	1020
GTGGATGGGG	CCGGCGGCCG	CCGCGATGGC	GGCCGCGGCCA	ACGCCGTATG	TGGGGTGGCT	1080
GGCCGCCACG	GGCCGCGCTGG	CGAAGGAGAC	GGCCACACAG	GCGAGGGCAG	GGCGGAAAGC	1140
GTGGATGGGG	CCGGCGGCCG	CCGCGATGGC	GGCCGCGGCCA	ACGCCGTATG	TGGGGTGGCT	1200
GGCCGCCACG	GGCCGCGCTGG	CGAAGGAGAC	GGCCACACAG	GCGAGGGCAG	GGCGGAAAGC	1260
GTGGATGGGG	CCGGCGGCCG	CCGCGATGGC	GGCCGCGGCCA	ACGCCGTATG	TGGGGTGGCT	1320
GGCCGCCACG	GGCCGCGCTGG	CGAAGGAGAC	GGCCACACAG	GCGAGGGCAG	GGCGGAAAGC	1380
GTGGATGGGG	CCGGCGGCCG	CCGCGATGGC	GGCCGCGGCCA	ACGCCGTATG	TGGGGTGGCT	1440
GG						1441

(2) INFORMATION FOR SEQ ID NO:142:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 423 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:142:

Met	Asp	Phe	Gly	Lle	Lle	Pro	Pro	Glu	Val	Asn	Ser	Ser	Arg	Met	Tyr
1				5					10				15		
Ser	Gly	Pro	Gly	Pro	Glu	Ser	Met	Lle	Ala	Ala	Ala	Ala	Trp	Asp	

20 25 30

Gly Val Ala Ala Glu Leu Thr Ser Ala Ala Val Ser Tyr Gly Ser Val
 35 40 45
 Val Ser Thr Leu Ile Val Glu Pro Trp Met Gly Pro Ala Ala Ala Ala
 50 55 60
 Met Ala Ala Ala Ala Thr Pro Tyr Val Gly Trp Leu Ala Ala Thr Ala
 65 70 75 80
 Ala Leu Ala Lys Glu Thr Ala Thr Gln Ala Arg Ala Ala Glu Ala
 85 90 95
 Phe Gly Thr Ala Phe Ala Met Thr Val Pro Pro Ser Leu Val Ala Ala
 100 105 110
 Asn Arg Ser Arg Leu Met Ser Leu Val Ala Ala Asn Ile Leu Gly Gln
 115 120 125
 Asn Ser Ala Ala Ile Ala Ala Thr Gln Ala Glu Tyr Ala Glu Met Trp
 130 135 140
 Ala Gln Asp Ala Ala Val Met Tyr Ser Tyr Glu Gly Ala Ser Ala Ala
 145 150 155 160
 Ala Ser Ala Leu Pro Pro Phe Thr Pro Pro Val Gln Gly Thr Gly Pro
 165 170 175
 Ala Gly Pro Ala Ala Ala Ala Ala Thr Gln Ala Ala Gly Ala Gly
 180 185 190
 Ala Val Ala Asp Ala Gln Ala Thr Leu Ala Gln Leu Pro Pro Gly Ile
 195 200 205
 Leu Ser Asp Ile Leu Ser Ala Leu Ala Ala Asn Ala Asp Pro Leu Thr
 210 215 220
 Ser Gly Leu Leu Gly Ile Ala Ser Thr Leu Asn Pro Gln Val Gly Ser
 225 230 235 240
 Ala Gln Pro Ile Val Ile Pro Thr Pro Ile Gly Glu Leu Asp Val Ile
 245 250 255
 Ala Leu Tyr Ile Ala Ser Ile Ala Thr Gly Ser Ile Ala Leu Ala Ile
 260 265 270
 Thr Asn Thr Ala Arg Pro Trp His Ile Gly Leu Tyr Gly Asn Ala Gly
 275 280 285
 Gly Leu Gly Pro Thr Gln Gly His Pro Leu Ser Ser Ala Thr Asp Glu
 290 295 300
 Pro Glu Pro His Trp Gly Pro Phe Gly Gly Ala Ala Pro Val Ser Ala
 305 310 315 320
 Gly Val Gly His Ala Ala Leu Val Gly Ala Leu Ser Val Pro His Ser
 325 330 335
 Trp Thr Thr Ala Ala Pro Glu Ile Gln Leu Ala Val Gln Ala Thr Pro
 340 345 350
 Thr Phe Ser Ser Ala Gly Ala Asp Pro Thr Ala Leu Asn Gly Met
 355 360 365
 Pro Ala Gly Leu Leu Ser Gly Met Ala Leu Ala Ser Leu Ala Ala Arg
 370 375 380
 Gly Thr Thr Gly Gly Gly Thr Arg Ser Gly Thr Ser Thr Asp Gly
 385 390 395 400
 Gln Glu Asp Gly Arg Lys Pro Pro Val Val Val Ile Arg Glu Gln Pro
 405 410 415
 Pro Pro Gly Asn Pro Pro Arg
 420

Mtb40 (HTCC#1)

(2) INFORMATION FOR SEQ ID NO:137:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1200 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:137:

CAGGCATGAG CAGAGCGTTC ATCATCGATC CAACGATCAG TGCCATTGAC GGCTTGTACG	60
ACCTTCTGGG GATTGGAATA CCCAACCAAG GGGGTATCCT TTACTCCTCA CTAGAGTACT	120
TCGAAAAAGC CCTGGAGGAG CTGGCAGCAG CGTTTCCGGG TGATGGCTGG TTAGGTTCGG	180
CCGGCGACAA ATACGCCGGC AAAAACCGCA ACCACGTGAA TTTTTCCAG GAACTGGCAG	240
ACCTCGATCG TCAGCTCATC AGCCTGATCC ACGACCAGGC CAACCGCGTC CAGACGACCC	300
GCGACATCCT GGAGGGCGCC AAGAAAGGTC TCGAGTTCGT GCGCCCGGTG GCTGTGGACC	360
TGACCTACAT CCCGGTCGTC GGGCACGCC TATCGGCCGC CTTCCAGGCG CCGTTTGCG	420
CGGGCGCGAT GGCGTAGTG GGCGGGCGC TTGCCTACTT GGTCGTGAAA ACGCTGATCA	480
ACGCGACTCA ACTCCTCAAA TTGCTTGCCA ATTGGCGGA GTTGGTCGCG GCCGCCATTG	540
CGGACATCAT TTCGGATGTG GCGGACATCA TCAAGGGCAC CCTCGGAGAA GTGTGGGAGT	600
TCATCACAAA CGCGCTCAAC GGCCTGAAAG AGCTTTGGGA CAAGCTCACG GGGTGGGTGA	660
CCGGACTGTT CTCTCGAGGG TGGTCAAACC TGGAGTCCTT CTTTGCAGGC GTCCCCGGCT	720
TGACCGGCGC GACCAGCGGC TTGTCGAAG TGACTGGCTT GTTCCGGCGC GCGGTCTGT	780
CCGCATCGTC GGGCTTGGCT CACCGGGATA GCCTGGCGAG CTCAGCCAGC TTGCCCCCCC	840
TGGCCGGCAT TGGGGCGGG TCCGGTTTG GGGGCTTGCC GAGCCTGGCT CAGGTCCATG	900
CCGCCTCAAC TCGGCAGGGC CTACGGCCCC GAGCTGATGG CCCGGTCGGC GCGCTGCCG	960
AGCAGGGTCGG CGGGCAGTCG CAGCTGGTCT CGCGCGAGGG TTCCCAAGGT ATGGGGCGGAC	1020
CCGTAGGCAT GGGCGGCATG CACCCCTCTT CGGGGGCGTC GAAAGGGACG ACGACGAAGA	1080
AGTACTCGGA AGCGCGGGCG CGGGGACTG AACACGCCGA GCGCGCGCCA GTCGAAGCTG	1140
ACGCGGGCGG TGGGCAAAAG GTGCTGGTAC GAAACGTCGT CTAACGGCAT GGCGAGCCAA	1200

(2) INFORMATION FOR SEQ ID NO:138:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 392 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:138:

Met Ser Arg Ala Phe Ile Ile Asp Pro Thr Ile Ser Ala Ile Asp Gly			
1	5	10	15
Leu Tyr Asp Leu Leu Gly Ile Gly Ile Pro Asn Gln Gly Gly Ile Leu			
20	25	30	
Tyr Ser Ser Leu Glu Tyr Phe Glu Lys Ala Leu Glu Glu Leu Ala Ala			
35	40	45	
Ala Phe Pro Gly Asp Gly Trp Leu Gly Ser Ala Ala Asp Lys Tyr Ala			
50	55	60	
Gly Lys Asn Arg Asn His Val Asn Phe Phe Gln Glu Leu Ala Asp Leu			
65	70	75	80
Asp Arg Gln Leu Ile Ser Leu Ile His Asp Gln Ala Asn Ala Val Gln			
85	90	95	
Thr Thr Arg Asp Ile Leu Glu Gly Ala Lys Lys Gly Leu Glu Phe Val			
100	105	110	

5
9
8
7
6
5
4
3
2
1

Arg Pro Val Ala Val Asp Leu Thr Tyr Ile Pro Val Val Gly His Ala
 115 120 125
 Leu Ser Ala Ala Phe Gln Ala Pro Phe Cys Ala Gly Ala Met Ala Val
 130 135 140
 Val Gly Gly Ala Leu Ala Tyr Leu Val Val Lys Thr Leu Ile Asn Ala
 145 150 155 160
 Thr Gln Leu Leu Lys Leu Leu Ala Lys Leu Ala Glu Leu Val Ala Ala
 165 170 175
 Ala Ile Ala Asp Ile Ile Ser Asp Val Ala Asp Ile Ile Lys Gly Thr
 180 185 190
 Leu Gly Glu Val Trp Glu Phe Ile Thr Asn Ala Leu Asn Gly Leu Lys
 195 200 205
 Glu Leu Trp Asp Lys Leu Thr Gly Trp Val Thr Gly Leu Phe Ser Arg
 210 215 220
 Gly Trp Ser Asn Leu Glu Ser Phe Phe Ala Gly Val Pro Gly Leu Thr
 225 230 235 240
 Gly Ala Thr Ser Gly Leu Ser Gln Val Thr Gly Leu Phe Gly Ala Ala
 245 250 255
 Gly Leu Ser Ala Ser Ser Gly Leu Ala His Ala Asp Ser Leu Ala Ser
 260 265 270
 Ser Ala Ser Leu Pro Ala Leu Ala Gly Ile Gly Gly Ser Gly Phe
 275 280 285
 Gly Gly Leu Pro Ser Leu Ala Gln Val His Ala Ala Ser Thr Arg Gln
 290 295 300
 Ala Leu Arg Pro Arg Ala Asp Gly Pro Val Gly Ala Ala Ala Glu Gln
 305 310 315 320
 Val Gly Gly Gln Ser Gln Leu Val Ser Ala Gln Gly Ser Gln Gly Met
 325 330 335
 Gly Gly Pro Val Gly Met Gly Gly Met His Pro Ser Ser Gly Ala Ser
 340 345 350
 Lys Gly Thr Thr Thr Lys Tyr Ser Glu Gly Ala Ala Gly Thr
 355 360 365
 Glu Asp Ala Glu Arg Ala Pro Val Glu Ala Asp Ala Gly Gly Gln
 370 375 380
 Lys Val Leu Val Arg Asn Val Val
 385 390

Mtb9.9A (MTI-A)

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1742 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CCGCTCTCTT TCAACGTCAT AAGTTGGTG GGCCAGTCGG CCGCGCGTGC ATATGGCACC
 AATAACCGGT GTCCCATGGA TACCCGGACC GCACGACGGT AGAGCGGATC AGCGCAGCCG

60

120

DRAFT DRAFT DRAFT

GTGCCGAACA	CTACCGCGTC	CACGCTCAGC	CCTGCCGCGT	TGCGGAAGAT	CGAGCCCAGG	180
TTCTCATGGT	CGTTAACGCC	TTCCAACACT	GCGACGGTGC	GCGCCCCGGC	GACCACCTGA	240
GCAACGCTCG	GCTCCGGCAC	CCGGCGCGCG	GCTGCCAACAA	CCCCACGATT	GAGATGGAAG	300
CCGATCACCC	GTGCCATGAC	ATCAGCCGAC	GCTCGATAGT	ACGGCGCGCC	GACACCGGCC	360
AGATCATCCT	TGAGCTCGGC	CAGCCGGCGG	TCGGTGCGGA	ACAGCGCCAG	CGCGTGAAC	420
CGTGAGGCCA	GCATGCGCTG	CACCACCGC	ACACCCCTCGG	CGATCACCAA	CGCCCTTGCGG	480
GTCGGCAGAT	CGGGACNACN	GTCGATGCTG	TTCAAGGTAC	GGAAATCGTC	GAGCCGTGGG	540
TCGTCGGGAT	CGCAGACGTC	CTGAACATCG	AGGCCGTCGG	GGTGCCTGGG	ACAACGGCCT	600
TCGGTCACGG	GCTTTCGTG	ACCAGAGCCA	GCATCAGATC	GGCGGCCGCTG	CGCAGGATGT	660
CACGCTCGCT	GCGGTTCA	GTCGCGAGCC	GCTCAGCCAG	CCACTCTTGC	AGAGAGCCGT	720
TGCTGGGATT	AATTGGGAGA	GGAAAGACAGC	ATGTCGTTCG	TGACCACACA	GCCGGAAGCC	780
CTGGCAGCTG	CGGCGGCCGAA	CCTACAGGGT	ATTGGCACGA	CAATGAACGC	CCAGAACGCG	840
GCCGCGCTG	CTCCAACCAC	CGGAGTAGTG	CCCGCAGCCG	CCGATGAAGT	ATCAGCGCTG	900
ACCGCGCTC	AGTTGCTGC	GCACGCGAG	ATGTACCAA	CGGTCA	CCAGGCCGCG	960
GCCATTACAG	AAATGTTCGT	GAACACGCTG	GTGGCCAGTT	CTGGCTCATA	CGCGGCCACC	1020
GAGGCGGCCA	ACGCAGCCG	TGCCGGCTGA	ACGGGCTCGC	ACGAACCTGC	TGAAGGAGAG	1080
GGGAAACATC	CGGAGTTCTC	GGGTCAGGGG	TTGCGCCAGC	GCCCAGCCGA	TTCAGNTATC	1140
GGCGTCCATA	ACAGCAGACG	ATCTAGGCAT	TCAGTACTAA	GGAGACAGGC	AACATGGCCT	1200
CACGTTTAT	GACGGATCCG	CATGCGATGC	GGGACATGGC	GGGCCGTTT	GAGGTGCACG	1260
CCCAGACGGT	GGAGGACGAG	GCTCGCCGGA	TGTGGGCGTC	CCCGCAAAAC	ATTTCGGTG	1320
CGGGCTGGAG	TGGCATGGCC	GAGGCACCT	CGCTAGACAC	CATGACCTAG	ATGAATCAGG	1380
CGTTTCGAA	CATCGTGAAC	ATGCTGACG	GGGTGCGTGA	CGGGCTGGT	CGCGACGCCA	1440
ACAANTACGA	ACAGCAAGAG	CAGGCCCTCC	AGCAGATCCT	GAGCAGNTAG	CGCCGAAAGC	1500
CACAGCTGNG	TACGNTTCT	CACATTAGGA	GAACACCAAT	ATGACGATTA	ATTACCAGTT	1560
CGGGGACGTC	GACGCTCATG	GCGCCATGAT	CCCGCCTCAG	GCGGCCTCGC	TTGAGGCGGA	1620
GCATCAGGCC	ATCGTTCGT	ATGTGTTGGC	CGCGGGTGAC	TTTTGGGGCG	GCGCCGGTTC	1680
GGTGGCTTGC	CAGGAGTTCA	TTACCCAGTT	GGGCCGTAAC	TTCCAGGTGA	TCTACGAGCA	1740
GG						1742

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2836 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: *Mycobacterium tuberculosis*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GTTGATTCCG	TTCGCGGCCG	CGCCGAAGAC	CACCAACTCC	GCTGGGGTGG	TCCGCACAGGC	60
GGTTGCGTCG	GTCAGCTGGC	CGAATCCAA	TGATTGGTGG	CTCNGTGC	GGTGTGGGCT	120
CGATTACCCC	CACGGAAAGG	ACGACGATCG	TTCTGTTGCT	CGGTCAGTCG	TACTTGGC	180
CGGGCATGGC	GCGGTTCTT	ACCTCGATCG	CACAGCAGCT	GACCTTCGGC	CCAGGGGGCA	240
CAACGGCTGG	CTCCGGCGGA	GCCTGGTAC	CAACGCCACA	ATTGCGCC	CTGGGTGCAG	300
GCCC GGCGGT	GTCGGCGAGT	TTGGCGCGGG	CGGAGCCGGT	CGGGAGGTG	TCGGTGC	360
CAAGTTGGC	CGTCGCGGCT	CCGGCCTTCG	CGGAGAAGCC	TGAGGCGGGC	ACGCCGATGT	420
CCGTCA	CGAAGCGTCC	AGCTGCGTC	AGGGAGGCCT	GCTTCGAGGC	ATACCGCTGG	480
CGAGAGCGGG	GCGCGTACA	GGCGCTT	CTCACCGATA	CGGGTCCGC	CACAGCGTGA	540
TTACCCGGTC	TCCGTCGGCG	GGATAGCTT	CGATCCGGTC	TGCGCGGCCG	CCGAAATGC	600
TGCA	GATCGACCGC	GGCGCGCGT	AAACGCCGCA	CACGGCACTA	TCAATGCGCA	660
CGGCGGGCGT	TGATGCCAA	TTGACCGTCC	CGACGGGGCT	TTATCTGCGG	CAAGATTTC	720
TCCCCAGCCC	GGTCGGTGGG	CCGATAAATA	CGCTGGTCAG	CGCGACTCTT	CCGGCTGAAT	780

DRAFT

TCGATGCTCT	GGGCGCCCG	TCGACGCCGA	GTATCTCGAG	TGGGCCGCAA	ACCCGGTCAA	840
ACGCTGTTAC	TGTGGCGTTA	CCACAGGTGA	ATTGCGGTG	CCAACTGGTG	AACACTTGC	900
AACGGGTGGC	ATCGAAATCA	ACTTGTGCG	TTGCA	GATG	CTACTCTCTT	960
GTTGCTGGGA	TTAATTGGGA	GAGGAAGACA	GCATGTCGTT	CGTGAC	CACA	1020
CCCTGGCAGC	TGCGGCGCG	AACCTACAGG	GTATTGGCAC	GACAATGAAC	GCCCAGAACG	1080
CGGCCGCGGC	TGCTCCAACC	ACCGGAGTAG	TGCGCGCAGC	CGCGATGAA	GTATCAGCGC	1140
TGACCGCGGC	TCA	AGATGTACCA	AACGGTCAGC	GCCCAGGCCG	1200	
CGGCCATTCA	CGAAATGTT	GTGAACACGC	TGCGGCCAG	TTCTGGCTCA	TACGCGGCCA	1260
CCGAGGCGGC	CAACGCAGCC	GCTGCCGCT	GAACGGGCTC	GCACGAACCT	GCTGAAGGAG	1320
AGGGGAAACA	TCCGGAGTT	TCGGGTCA	GGTTGCGCCA	GCGCCCAGCC	GATTCA	1380
TCGGCGTCCA	TAACAGCAGA	CGATCTAGC	ATTCA	AAGGAGACAG	GCAACATGGC	1440
CTCACGTTT	ATGACGGATC	CGCATGCGAT	GCGGACATG	GCGGCCGTT	TTGAGGTGCA	1500
CGCCCAGACG	TGGGAGGAGC	AGGCTGCCG	GATGTTGGCG	TCCGCGAAA	ACATTTCGG	1560
TGCGGGCTGG	AGTGGCATGG	CCGAGGCCAC	CTCGCTAGAC	ACCATGACCT	AGATGAATCA	1620
GGCGTTTCG	AACATCGTGA	ACATGTCGA	CGGGGTGCGT	GACGGGCTGG	TTCGCGACGC	1680
CAACAACTAC	GAACAGCAAG	AGCAGGCC	CCAGCAGATC	CTGAGCAGCT	AGCGCCGAAA	1740
GCCACAGCTG	CGTACGCTTT	CTCACATTAG	GAGAACACCA	ATATGACGAT	TAATTACCA	1800
TTCGGGGAGC	TCGACGCTCA	TGGCGCCATG	ATCCGCGCTC	AGGCGGGCTC	GCTTGAGGCG	1860
GAGCATCAGG	CCATCGTTCG	TGATGTGTTG	GCCGCGGGTG	ACTTTGGGG	CGCGCCGGT	1920
TCGGTGGCTT	GCCAGGAGTT	CATTACCCAG	TTGGGCCGTA	ACTTCCAGGT	GATCTACGAG	1980
CAGGCCAACG	CCCACGGCA	GAAGGTGCA	GCTGCCGCA	ACAACATGGC	GCAAACCGAC	2040
AGCGCCGTCG	GCTCCAGCTG	GGCCTAAAAC	TGAAC	TCGCGGCAGC	ACACCAACCA	2100
GCGGTGTG	TGCTGTGTG	TGCA	TAGCACTCGA	CCGCTGAGGT	AGCGATGGAT	2160
CAACAGAGTA	CCCGCACCGA	CATCACCGTC	AACGTCGACG	GCTTCTGGAT	GCTTCAGGCG	2220
CTACTGGATA	TCCGCCACGT	TGCGCTGAG	TTACGTTG	GGCGTACGT	CTCCACCGAT	2280
TCCAATGACT	GGCTAAACGA	GCACCCGGG	ATGGCGGTCA	TGCGCGAGCA	GGCATTGTC	2340
GTCAACGACG	CGGTCAACGA	ACAGGTCGCT	GCCCCGATGA	AGGTGCTTGC	CGCACCTGAT	2400
CTTGAAGTCG	TCGCCCTGCT	GTCACGCGC	AAAGTTGCTGT	ACGGGGTCAT	AGACGACGAG	2460
AACCAGCCGC	CGGGTTCGCG	TGACATCCCT	GACAATGAGT	TCCGGGTGGT	GTTGGCCCGG	2520
CGAGGCCAGC	ACTGGGTGTC	GGCGGTACGG	GTTGGCAATG	ACATCACCGT	CGATGACGTG	2580
ACGGTCTCGG	ATAGCGCTC	GATCGCCGA	CTGGTAATGG	ACGGTCTGG	GTCGATTAC	2640
CACGCCGACC	CAGCCCGCAT	CAACGCGTC	AACGTGCCAA	TGGAGGAGAT	CTCGTGCCGA	2700
ATTCCGGCAGC	AGGCACGAGG	CGGTGTCGGT	GACGACGGGA	TCGATCACGA	TCATCGACCG	2760
GCCGGGATCC	TTGGCGATCT	CGTTGAGCAC	GACCCGGGCC	CGCGGGAAAGC	TCTCGACAT	2820
CCATGGGTT	TTCCCG					2836

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 94 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii). MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Mycobacterium tuberculosis*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Met	Thr	Ile	Asn	Tyr	Gln	Phe	Gly	Asp	Val	Asp	Ala	His	Gly	Ala	Met
1						5				10					15
Ile	Arg	Ala	Leu	Ala	Gly	Leu	Leu	Glu	Ala	Glu	His	Gln	Ala	Ile	Ile
						20				25					30
Ser	Asp	Val	Leu	Thr	Ala	Ser	Asp	Phe	Trp	Gly	Gly	Ala	Gly	Ser	Ala
						35				40					45

095E3363672 21094000

Ala Cys Gln Gly Phe Ile Thr Gln Leu Gly Arg Asn Phe Gln Val Ile
50 55 60
Tyr Glu Gln Ala Asn Ala His Gly Gln Lys Val Gln Ala Ala Gly Asn
65 70 75 80
Asn Met Ala Gln Thr Asp Ser Ala Val Gly Ser Ser Trp Ala
85 90

Mtb9.9A (MTI-A) ORF peptides

(2) INFORMATION FOR SEQ ID NO:51:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

Met Thr Ile Asn Tyr Gln Phe Gly Asp Val Asp Ala His Gly Ala
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:52:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

Gln Phe Gly Asp Val Asp Ala His Gly Ala Met Ile Arg Ala Gln
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:53:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:

DRAFT

(A) ORGANISM: *Mycobacterium tuberculosis*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

Asp Ala His Gly Ala Met Ile Arg Ala Gln Ala Ala Ser Leu Glu
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Mycobacterium tuberculosis*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

Met Ile Arg Ala Gln Ala Ala Ser Leu Glu Ala Glu His Gln Ala
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Mycobacterium tuberculosis*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

Ala Ala Ser Leu Glu Ala Glu His Gln Ala Ile Val Arg Asp Val
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Mycobacterium tuberculosis*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

Ala Glu His Gln Ala Ile Val Arg Asp Val Leu Ala Ala Gly Asp
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:57:

- (i) SEQUENCE CHARACTERISTICS:

 - (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
- (A) ORGANISM: *Mycobacterium tuberculosis*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

Ile Val Arg Asp Val Leu Ala Ala Gly Asp Phe Trp Gly Gly Ala
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:58:

- (i) SEQUENCE CHARACTERISTICS:

 - (A) LENGTH: 16 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: *Mycobacterium tuberculosis*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

(2) INFORMATION FOR SEQ ID NO:59:

- (i) SEQUENCE CHARACTERISTICS:

 - (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: *Mycobacterium tuberculosis*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

(2) INFORMATION FOR SEQ ID NO:60:

DOCUMENT NUMBER

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

Gly Ser Val Ala Cys Gln Glu Phe Ile Thr Gln Leu Gly Arg Asn
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:61:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

Gln Glu Phe Ile Thr Gln Leu Gly Arg Asn Phe Gln Val Ile Tyr Glu
1 5 10 15
Gln Ala

(2) INFORMATION FOR SEQ ID NO:62:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

Arg Asn Phe Gln Val Ile Tyr Glu Gln Ala Asn Ala His Gly Gln
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:63:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids

DECODED

- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

Ile Tyr Glu Gln Ala Asn Ala His Gly Gln Lys Val Gln Ala Ala
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:64:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

Asn Ala His Gly Gln Lys Val Gln Ala Ala Gly Asn Asn Met Ala
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:65:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

Lys Val Gln Ala Ala Gly Asn Asn Met Ala Gln Thr Asp Ser Ala
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:66:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 16 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

Gly Asn Asn Met Ala Gln Thr Asp Ser Ala Val Gly Ser Ser Trp Ala
1 5 10 15

Mtb9.8 (MSL)

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 585 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

TGGATTCCGA TAGCGGTTTC GGCCCCCTCGA CGGGCGACCA CGGGCGCGAG GCCTCCGAAC	60
GGGGGGCCGG GACGCTGGGA TTGCGCCGGGA CCCCAACCAA AGAACGCCGG GTCCGGGCGG	120
TCGGGCTGAC CGCACTGGCC GGTGATGAGT TCGGCAACGG CCCCCGGATG CCGATGGTGC	180
CGGGGACCTG GGAGCAGGGC AGCAACGAGC CCCAGGGCGCC CGACGGATCG GGGAGAGGGG	240
GAGGCACGG CTTACCGCAC GACAGCAAGT AACCGAATTG CGAACATCACGT GGACCCGTAC	300
GGGTCGAAAG GAGAGATGTT ATGAGCCTT TGGATGCTCA TATCCCACAG TTGGTGGCCT	360
CCCAGTCGGC GTTTGCGGCC AAGGCAGGGC TGATGCGGCA CACGATCGGT CAGGCGAGC	420
AGGCAGGGAT GTCGGCTCA GCGTTTCACC AGGGGGAGTC GTCGGCGGCG TTTCAGGCCG	480
CCCATGCCCG GTTGTGGCG GCGGCCGCCA AAGTCAACAC CTTGTTGGAT GTCGCGCAGG	540
CGAATCTGGG TGAGGCCGCC GGTACCTATG TGGCCGCCGA TGCTG	585

(2) INFORMATION FOR SEQ ID NO:109:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 97 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:109:

Met Ser Leu Leu Asp Ala His Ile Pro Gln Leu Val Ala Ser Gln Ser
1 5 10 15
Ala Phe Ala Ala Lys Ala Gly Leu Met Arg His Thr Ile Gly Gln Ala
20 25 30
Glu Gln Ala Ala Met Ser Ala Gln Ala Phe His Gln Gly Glu Ser Ser

35 40 45
Ala Ala Phe Gln Ala Ala His Ala Arg Phe Val Ala Ala Ala Ala Lys
50 55 60
Val Asn Thr Leu Leu Asp Val Ala Gln Ala Asn Leu Gly Glu Ala Ala
65 70 75 80
Gly Thr Tyr Val Ala Ala Asp Ala Ala Ala Ser Thr Tyr Thr Gly
85 90 95
Phe

Mtb9.8 ORF peptides

(2) INFORMATION FOR SEQ ID NO:110:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:110:

Met Ser Leu Leu Asp Ala His Ile Pro Gln Leu Val Ala Ser Gln
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:111:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:111:

Ala His Ile Pro Gln Leu Val Ala Ser Gln Ser Ala Phe Ala Ala
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:112:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:112:

Leu Val Ala Ser Gln Ser Ala Phe Ala Ala Lys Ala Gly Leu Met
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:113:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:113:

Ser Ala Phe Ala Ala Lys Ala Gly Leu Met Arg His Thr Ile Gly
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:114:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:114:

Lys Ala Gly Leu Met Arg His Thr Ile Gly Gln Ala Glu Gln Ala
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:115:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:115:

Arg His Thr Ile Gly Gln Ala Glu Gln Ala Ala Met Ser Ala Gln
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:116:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:116:

Gln Ala Glu Gln Ala Ala Met Ser Ala Gln Ala Phe His Gln Gly

1

5

10

15

(2) INFORMATION FOR SEQ ID NO:117:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:117:

Ala Met Ser Ala Gln Ala Phe His Gln Gly Glu Ser Ser Ala Ala
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:118:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:118:

Ala Phe His Gln Gly Glu Ser Ser Ala Ala Phe Gln Ala Ala His
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:119:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:119:

Glu Ser Ser Ala Ala Phe Gln Ala Ala His Ala Arg Phe Val Ala
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:120:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:120:

Phe Gln Ala Ala His Ala Arg Phe Val Ala Ala Ala Lys Val
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:121:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:121:

Ala Arg Phe Val Ala Ala Ala Lys Val Asn Thr Leu Leu Asp
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:122:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:122:

Ala Ala Ala Lys Val Asn Thr Leu Leu Asp Val Ala Gln Ala Asn
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:123:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:123:

Asn Thr Leu Leu Asp Val Ala Gln Ala Asn Leu Gly Glu Ala Ala
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:124:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:124:

Val Ala Gln Ala Asn Leu Gly Glu Ala Ala Gly Thr Tyr Val Ala Ala
1 5 10 15
Asp Ala

Mtb39A (TbH9)

(2) INFORMATION FOR SEQ ID NO:106:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3058 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:106:

GATCGTACCC	GTGCGAGTGC	TGGGGCCGTT	TGAGGATGGA	GTGCACGTGT	CTTCGTGAT	60
GGCATACCCA	GAGATGTTGG	CGGCGCCGGC	TGACACCCCTG	CAGAGCATCG	GTGCTACCAC	120
TGTGGCTAGC	AATGCCGCTG	CGGCGGGCCC	GACGACTGGG	GTGGTCCCCC	CCGCTGCCGA	180
TGAGGTGTCG	GCGCTGACTG	CGGCGCACAT	GCAGCGATGT	ATCAGTCCGT	240	
GAGCGCTCGG	GCTGCTGCGA	TTCATGACCA	GTTCGTGGCC	ACCCCTGCCA	GCAGCGCCAG	300
CTCGTATGCG	GCCACTGAAG	TCGCCAAATGC	GGCGGCGGGC	AGCTAACCCA	GGAACAGTCG	360
GCACGAGAAA	CCACCGAGAAA	TAGGGACACG	TAATGGTGG	TTTCGGGGCG	TTACCAACCGG	420
AGATCAACTC	CGCGAGGATG	TACGCCGGCC	CGGGTTCGGC	CTCGCTGGTG	GCCGCGGCTC	480
AGATGTGGGA	CAGCGTGGCG	AGTGACCTGT	TTTCGGCCGC	GTCGGCGTTT	CAGTCGGTGG	540
TCTGGGTCT	GACGGTGGGG	TCGTGGATAG	GTTCGTGGC	GGGTCTGATG	GTGGCGGGCG	600
CCTCGCCGTA	TGTGGCGTGG	ATGAGCGTCA	CCGCGGGGCA	GGCGAGCTG	ACCGCCGCCC	660
AGGTCCGGGT	TGCTGCGGCG	GCCTACGAGA	CGGCGTATGG	GCTGACGGTG	CCCCCGCCGG	720
TGATGCCGGA	GAACCGTGCT	GAACGTATGA	TTCTGATAGC	GACCAAACCTC	TTGGGGCAAA	780
ACACCCCGGC	GATCGCGGTC	AACGAGGCCG	AATACGGCGA	GATGTGGGCC	CAAGACGCCG	840
CCCGGATGTT	TGGCTACGCC	GCGCGACCGG	CGACGGCGAC	GGCGACGTTG	CTGCCGTTCG	900
AGGAGGCGCC	GGAGATGACC	AGCGCGGGTG	GGCTCCTCGA	GCAGGCCGCC	GCGGTGAGG	960
AGGCCTCCGA	CACCGCCGCG	GCGAACCAAGT	TGATGAACAA	TGTCCCCAG	GCGCTGCAAC	1020
AGCTGGCCA	GCCCACCGCAG	GGCACCAACGC	CTTCTTCCAA	GCTGGGTGGC	CTGTGGAAGA	1080
CGGTCTCGCC	GCATCGGTG	CCGATCAGCA	ACATGGTGTG	GATGGCCAAC	AACCACATGT	1140
CGATGACCAA	CTCGGGTGTG	TCGATGACCA	ACACCTTGAG	CTCGATGTTG	AAGGGCTTG	1200
CTCCGGCGGC	GGCCGCCAG	GCGTGCAGA	CCGCGGCGCA	AAACGGGTC	CGGGCGATGA	1260
GCTCGCTGGG	CAGCTCGCTG	GGTTCTTCGG	GTCTGGCGG	TGGGGTGGCC	GCCAACTTGG	1320
GTCGGCGGC	CTCGGTGGT	TGTTGTCGG	TGCCGAGGC	CTGGGCCGCG	GCCAACCAGG	1380
CAGTCACCCC	GGCGGCCGGG	GCGCTGCCG	TGACCAAGCT	GACCAACGCC	GCGAAAGAG	1440
GGCCCGGGCA	GATGCTGGGC	GGGCTGCCG	TGGGGCAGAT	GGGCGCCAGG	GCCGGTGGTG	1500
GGCTCAGTGG	TGTGCTGCGT	GTTCCGCCG	GACCTATGT	GATGCCGCA	TCTCCGGCGG	1560
CCGGCTAGGA	GAGGGGGCGC	AGACTGTCGT	TATTGACCA	GTGATCGCG	GTCTCGGTGT	1620
TTCCCGGGCC	GGCTATGACA	ACAGTCAATG	TGCAATGACAA	GTTACAGGTA	TTAGGTCCAG	1680
GTTCAACAAAG	GAGACAGGCC	ACATGGCCTC	ACGTTTATG	ACGGATCCG	ACGGCATGCC	1740
GGACATGGCG	GGCCGTTTG	AGGTGCACGC	CCAGACGGTG	GAGGACGAGG	CTCGCCGGAT	1800
GTGGGCGTCC	GCGCAAAACA	TTTCCGGTGC	GGGCTGGAGT	GGCATGCCG	AGGCGACCTC	1860
GCTAGACACC	ATGGCCCAGA	TGAATCAGGC	GTTTCGCAAC	ATCGTGAACA	TGCTGCACGG	1920
GGTGCCTGAC	GGGCTGGTTC	GCGACGCCA	CAACTACGAG	CAGCAAGAGC	AGGCCTCCCA	1980
GCAGATCCTC	AGCAGCTAAC	GTCAGCCGCT	GCAGCACAA	ACTTTTACAA	GCGAAGGAGA	2040
ACAGGTTCGA	TGACCATCAA	CTATCAATTG	GGGGATGTG	ACGCTCACGG	CGCCATGATC	2100
CGCGCTCAGG	CGGGGTTGCT	GGAGGCCGAG	CATCAGGCCA	TCATTCGTGA	TGTGTTGACC	2160
CGCAGTGACT	TTTGGGGCGG	CGCCGGTTCG	GGGGGTTCAT	TACCCAGTTG		2220

GGCCGTAACT	TCCAGGTGAT	CTACGAGCAG	GCCAACGCC	ACGGGCAGAA	GGTGCAGGCT	2280
GCCGGCAACA	ACATGGCGA	AACCGACAGC	GCCGTCGGCT	CCAGCTGGC	CTGACACCAG	2340
GCCAAGGCCA	GGGACGTGGT	GTACGAGTGA	AGTCCTCGC	GTGATCCCTC	GGGTGGCAGT	2400
CTAAGTGGTC	AGTGCTGGGG	TGTTGGTGGT	TTGCTGCTTG	GCGGGTTCTT	CGGTGCTGGT	2460
CAGTGCTGCT	CGGGCTCGGG	TGAGGACCTC	GAGGCCAGG	TAGCGCCGTC	CTTCGATCCA	2520
TTCGTCGTGT	TGTCGGCGA	GGACGGCTCC	GACGGAGCGG	ATGATCGAGG	CGCGGTGGG	2580
GAAGATGCC	ACGACGTGCG	TTCGGCGTCG	TACCTCTCGG	TTGAGGCAGT	CCTGGGGGTT	2640
GTTGGACCAG	ATTGGCGCC	AGATCTGCTT	GGGAAAGGCG	GTGAACGCCA	GCAGGTCGGT	2700
GCGGGCGGTG	TCGAGGTGCT	CGGCCACCGC	GGGGAGTTG	TCGGTCAGAG	CGTCGAGTAC	2760
CCGATCATAT	TGGCAACAA	CTGATTGCGC	GTCGGGCTGG	TCGTAGATGG	AGTGCAGCAG	2820
GGTGCACCC	CACGGCCAGG	AGGGCTTCGG	GGTGGCTGCC	ATCAGATTGG	CTGCGTAGTG	2880
GGTTCTGCAG	CGCTGCCAGG	CCGCTGGGG	CAGGGTGGCG	CCGATCGCGG	CCACCAAGGCC	2940
GGCGTGGCG	TCGCTGGTGA	CCAGCGCAC	CCCCGACAGG	CCGCAGGCCA	CCAGGTGCGC	3000
GAAGAACGCC	AGCCAGCCGG	CCCCGTCC	GGCGGAGGTG	ACCTGGATGC	CCAGGATC	3058

(2) INFORMATION FOR SEQ ID NO:107:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 391 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:

Met	Val	Asp	Phe	Gly	Ala	Leu	Pro	Pro	Glu	Ile	Asn	Ser	Ala	Arg	Met
1						5				10					15

Tyr	Ala	Gly	Pro	Gly	Ser	Ala	Ser	Leu	Val	Ala	Ala	Ala	Gln	Met	Trp
								20		25				30	

Asp	Ser	Val	Ala	Ser	Asp	Leu	Phe	Ser	Ala	Ala	Ser	Ala	Phe	Gln	Ser
						35		40						45	

Val	Val	Trp	Gly	Leu	Thr	Val	Gly	Ser	Trp	Ile	Gly	Ser	Ser	Ala	Gly
						50		55		60					

Leu	Met	Val	Ala	Ala	Ala	Ser	Pro	Tyr	Val	Ala	Trp	Met	Ser	Val	Thr
	65						70			75				80	

Ala	Gly	Gln	Ala	Glu	Leu	Thr	Ala	Ala	Gln	Val	Arg	Val	Ala	Ala	Ala
						85			90					95	

Ala	Tyr	Glu	Thr	Ala	Tyr	Gly	Leu	Thr	Val	Pro	Pro	Pro	Val	Ile	Ala
						100		105						110	

Glu	Asn	Arg	Ala	Glu	Leu	Met	Ile	Leu	Ile	Ala	Thr	Asn	Leu	Leu	Gly
						115		120						125	

Gln	Asn	Thr	Pro	Ala	Ile	Ala	Val	Asn	Glu	Ala	Glu	Tyr	Gly	Glu	Met
							130		135					140	

Trp	Ala	Gln	Asp	Ala	Ala	Ala	Met	Phe	Gly	Tyr	Ala	Ala	Ala	Thr	Ala
							145		150		155			160	

Thr	Ala	Thr	Ala	Thr	Leu	Leu	Pro	Phe	Glu	Ala	Pro	Glu	Met	Thr	
							165		170					175	

DRAFT

Ser	Ala	Gly	Gly	Leu	Leu	Glu	Gln	Ala	Ala	Ala	Val	Glu	Glu	Ala	Ser
180						185						190			
Asp	Thr	Ala	Ala	Ala	Asn	Gln	Leu	Met	Asn	Asn	Val	Pro	Gln	Ala	Leu
195							200					205			
Gln	Gln	Leu	Ala	Gln	Pro	Thr	Gln	Gly	Thr	Thr	Pro	Ser	Ser	Lys	Leu
210							215					220			
Gly	Gly	Leu	Trp	Lys	Thr	Val	Ser	Pro	His	Arg	Ser	Pro	Ile	Ser	Asn
225						230				235			240		
Met	Val	Ser	Met	Ala	Asn	Asn	His	Met	Ser	Met	Thr	Asn	Ser	Gly	Val
	245							250					255		
Ser	Met	Thr	Asn	Thr	Leu	Ser	Ser	Met	Leu	Lys	Gly	Phe	Ala	Pro	Ala
	260							265					270		
Ala	Ala	Ala	Gln	Ala	Val	Gln	Thr	Ala	Ala	Gln	Asn	Gly	Val	Arg	Ala
	275						280					285			
Met	Ser	Ser	Leu	Gly	Ser	Ser	Leu	Gly	Ser	Ser	Gly	Leu	Gly	Gly	Gly
	290						295					300			
Val	Ala	Ala	Asn	Leu	Gly	Arg	Ala	Ala	Ser	Val	Gly	Ser	Leu	Ser	Val
	305					310				315			320		
Pro	Gln	Ala	Trp	Ala	Ala	Ala	Asn	Gln	Ala	Val	Thr	Pro	Ala	Ala	Arg
	325							330					335		
Ala	Leu	Pro	Leu	Thr	Ser	Leu	Thr	Ser	Ala	Ala	Glu	Arg	Gly	Pro	Gly
	340							345					350		
Gln	Met	Leu	Gly	Gly	Leu	Pro	Val	Gly	Gln	Met	Gly	Ala	Arg	Ala	Gly
	355							360					365		
Gly	Gly	Leu	Ser	Gly	Val	Leu	Arg	Val	Pro	Pro	Arg	Pro	Tyr	Val	Met
	370						375					380			
Pro	His	Ser	Pro	Ala	Ala	Gly									
	385					390									

Mtb32A (TbRa35)

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1872 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GACTACGTTG GTGTAGAAAA ATCCTGCCGC CGCGACCCTT AAGGCTGGGA CAATTTCTGA

60

DRAFT DRAFT DRAFT

TAGCTACCCC	GACACAGGAG	GTTACGGGAT	GAGCAATTG	CGCCGCCGCT	CACTCAGGTG	120
GTCATGGTTG	CTGAGCGTGC	TGGCTGCCGT	CGGGCTGGC	CTGGCCACGG	CGCCGGCCCA	180
GGCGGGCCCCG	CCGGCCTTGT	CGCAGGACCG	GTTCGCCGAC	TTCCCCCGC	TGCCCCCTCGA	240
CCC GTCCCGC	ATGGTCGCC	AAGTGGCGCC	ACAGGTGGTC	AACATCAACA	CCAAACGTGG	300
CTACAACAAAC	GCCGTGGCG	CCGGGACCGG	CATCGTCATC	GATCCCAACG	GTGTCGTGCT	360
GACCAACAAAC	CACGTGATCG	CGGGCGCCAC	CGACATCAAT	GCGTTCAGCG	TCGGCTCCGG	420
CCAAACCTAC	GGCGTCGATG	TGGTCGGGTA	TGACCGCACC	CAGGATGTCG	CGGTGCTGCA	480
GCTCGCGG	GGCGGTGGCC	TGCCGTGGC	GGCGATCGGT	GGCGGCCGCG	CGGGTGGTGA	540
GCCCGTCGTC	CGCATGGGCA	ACAGCGGTGG	GCAGGGCGGA	ACGCCCGTG	CGGTGCCCTGG	600
CAGGGTGGTC	CGCCTCGGCC	AAACCGTGCA	GGCGTCGGAT	TCGCTGACCG	GTGCCGAAGA	660
GACATTGAAC	GGGTTGATCC	AGTTGATGC	CGCAATCCAG	CCCGGTGATT	CGGGCGGGCC	720
CGTCGTCAAC	GGCCTAGGAC	AGGTGGTCG	TATGAACACG	GCCGCGTCCG	ATAACTTCCA	780
GCTGTCCAG	GGTGGGCAGG	GATTGCCAT	TCCGATCGGG	CAGGCGATGG	CGATCGCGGG	840
CCAAATCCGA	TCGGGTGGGG	GGTCACCCAC	CGTTCATATC	GGGCCTACCG	CCTTCCTCGG	900
CTTGGGTGTT	GTGACAACA	ACGGCAACGG	CGCACAGTC	CAACCGGTGG	TCGGAAGCGC	960
TCCGGCGGCA	AGTCTCGGCA	TCTCACCGG	CGACGTGATC	ACCGCGGTCG	ACGGCGCTCC	1020
GATCAACTCG	GCCACCGCGA	TGGCGACGC	GCTTAACGGG	CATCATCCCG	GTGACGTCAT	1080
CTCGGTGAAC	TGGCAAACCA	AGTCGGGCGG	CACCGTACA	GGGAACGTGA	CATTGGCCGA	1140
GGGACCCCCG	GCCTGATTTG	TCGCGGATAC	CACCCGCC	CCGGCCAATT	GGATTGGCGC	1200
CAGCGTGT	TGCCCGGTGA	GCCCCCGAGT	TCCGCTCTCC	GTGCGCGTGG	CATTGTGGAA	1260
GCAATGAACG	AGGCAGAACAA	CAGCGTTGAG	CACCCCTCCCG	TGCAGGGCAG	TTACGTCGAA	1320
GGCGGTGTGG	TCGAGCATCC	GGATGCCAAG	GACTTCGGCA	GCGCCGCC	CCTGCCCGCC	1380
GATCCGACCT	GGTTTAAGCA	CGCCGTCTTC	TACCGAGGTGC	TGGTCCGGC	GTTCTTCGAC	1440
GCCAGCGCGG	ACGTTCCGN	CGATCTCGT	GGACTCATCG	ATCGCCTCGA	CTACCTGCAG	1500
TGGCTGGCA	TCGACTGCAT	CTGTTGCCG	CGTTCTACG	ACTCACCGCT	GCGCGACGGC	1560
GGTTACGACA	TTCGCGACTT	CTACAAGGTG	CTGCCGAAT	TCGGCACCGT	CGACGATTTC	1620
GTCGCCCCGG	TCGACACCGC	TCACCGGCGA	GGTATCCGCA	TCATCACCGA	CCTGGTGATG	1680
AATCACACCT	CGGAGTCGCA	CCCCTGGTTT	CAGGAGTC	GCGCGACCC	AGACGGACCG	1740
TACGGTGACT	ATTACGTGTG	GAGCGACACC	AGCGAGCGCT	ACACCGACGC	CCGGATCATC	1800
TTCGTCGACA	CCGAAGAGTC	GAACCTGGTCA	TTCGATCCTG	TCCGCCGACA	GTTNCTACTG	1860
GCACCGATT	TT					1872

(2) INFORMATION FOR SEQ ID NO:79:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 355 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

Met	Ser	Asn	Ser	Arg	Arg	Ser	Lau	Arg	Trp	Ser	Trp	Lau	Lau	Ser	
1				5				10				15			
Val	Lau	Ala	Ala	Val	Gly	Lau	Gly	Lau	Ala	Thr	Ala	Pro	Ala	Gln	Ala
								20		25			30		
Ala	Pro	Pro	Ala	Lau	Ser	Gln	Asp	Arg	Phe	Ala	Asp	Phe	Pro	Ala	Lau
							35	40		45					
Pro	Lau	Asp	Pro	Ser	Ala	Met	Val	Ala	Gln	Val	Ala	Pro	Gln	Val	Val
						50		55		60					
Asn	Ile	Asn	Thr	Lys	Lau	Gly	Tyr	Asn	Asn	Ala	Val	Gly	Ala	Gly	Thr
						65		70		75			80		
Gly	Ile	Val	Ile	Asp	Pro	Asn	Gly	Val	Val	Lau	Thr	Asn	Asn	His	Val
						85		90			95				
Ile	Ala	Gly	Ala	Thr	Asp	Ile	Asn	Ala	Phe	Ser	Val	Gly	Ser	Gly	Gln
						100		105			110				
Thr	Tyr	Gly	Val	Asp	Val	Val	Gly	Tyr	Asp	Arg	Thr	Gln	Asp	Val	Ala

BIOSEQUENCES

115	120	125
Val Leu Gln Leu Arg Gly Ala Gly Gly Leu Pro Ser Ala Ala Ile Gly		
130	135	140
Gly Gly Val Ala Val Gly Glu Pro Val Val Ala Met Gly Asn Ser Gly		
145	150	155
Gly Gln Gly Gly Thr Pro Arg Ala Val Pro Gly Arg Val Val Ala Leu		
165	170	175
Gly Gln Thr Val Gln Ala Ser Asp Ser Leu Thr Gly Ala Glu Glu Thr		
180	185	190
Leu Asn Gly Leu Ile Gln Phe Asp Ala Ala Ile Gln Pro Gly Asp Ser		
195	200	205
Gly Gly Pro Val Val Asn Gly Leu Gly Gln Val Val Gly Met Asn Thr		
210	215	220
Ala Ala Ser Asp Asn Phe Gln Leu Ser Gln Gly Gly Gln Gly Phe Ala		
225	230	235
Ile Pro Ile Gly Gln Ala Met Ala Ile Ala Gly Gln Ile Arg Ser Gly		
245	250	255
Gly Gly Ser Pro Thr Val His Ile Gly Pro Thr Ala Phe Leu Gly Leu		
260	265	270
Gly Val Val Asp Asn Asn Gly Asn Gly Ala Arg Val Gln Arg Val Val		
275	280	285
Gly Ser Ala Pro Ala Ala Ser Leu Gly Ile Ser Thr Gly Asp Val Ile		
290	295	300
Thr Ala Val Asp Gly Ala Pro Ile Asn Ser Ala Thr Ala Met Ala Asp		
305	310	315
Ala Leu Asn Gly His His Pro Gly Asp Val Ile Ser Val Asn Trp Gln		
325	330	335
Thr Lys Ser Gly Gly Thr Arg Thr Gly Asn Val Thr Leu Ala Glu Gly		
340	345	350
Pro Pro Ala		
355		

Mtb8.4 (DPV)

(2) INFORMATION FOR SEQ ID NO:101:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 500 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:

CGTGGCAATG TCGTTGACCG TCAGGGCCGG GGTGCCTCC GCAGATCCCG TGGACGCGGT	60
CATTAACACC ACCTGCAATT ACGGGGCAGGT AGTAGCTGCG CTCAACGCGA CGGATCCGGG	120
GGCTGCCGCA CAGTTAACG CCTCACCGGT GGCGCAGTCC TATTTGCGCA ATTCCTCGC	180
CGCACCGCCA CCTCAGCGCG CTGCGATGGC CGCGCAATTG CAAGCTGTGC CGGGGGCGGC	240
ACAGTACATC GGCCTTGTGAGTCGGTTGC CGGCTCCTGC AACAACTATT AAGCCCCATGC	300
GGGGCCCACATC CGCGGACCCG GCATCGTCGC CGGGGCTAGG CCAGATTGCC CCGCTCCTCA	360
ACGGGGCCGCA TCCCGCGACC CGGCATCGTC GCCGGGGCTA GGCCAGATTG CCCCGCTCCT	420
CAACGGGGCG CATCTCGTGC CGAATTCTTG CAGCCCCGGG GATCCACTAG TTCTAGAGCG	480
GCCGCCACCG CGGTGGAGCT	500

(2) INFORMATION FOR SEQ ID NO:102:

DRAFT DRAFT DRAFT

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 96 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:

Val	Ala	Met	Ser	Leu	Thr	Val	Gly	Ala	Gly	Val	Ala	Ser	Ala	Asp	Pro
1				5				10						15	
Val	Asp	Ala	Val	Ile	Asn	Thr	Thr	Cys	Asn	Tyr	Gly	Gln	Val	Val	Ala
				20				25						30	
Ala	Leu	Asn	Ala	Thr	Asp	Pro	Gly	Ala	Ala	Ala	Gln	Phe	Asn	Ala	Ser
	35					40						45			
Pro	Val	Ala	Gln	Ser	Tyr	Leu	Arg	Asn	Phe	Leu	Ala	Ala	Pro	Pro	Pro
	50					55						60			
Gln	Arg	Ala	Ala	Met	Ala	Ala	Gln	Leu	Gln	Ala	Val	Pro	Gly	Ala	Ala
65					70					75			80		
Gln	Tyr	Ile	Gly	Leu	Val	Glu	Ser	Val	Ala	Gly	Ser	Cys	Asn	Asn	Tyr
					85				90				95		

Mtb11 (Tb38-1)

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 327 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

CGGCACGAGA	GACCGATGCC	GCTACCCTCG	CGCAGGAGGC	AGGTAATTTC	GAGCGGATCT	60
CCGGCGACCT	GAAAACCCAG	ATCGACCAAG	TGGAGTCGAC	GGCAGGTTCG	TTGCAGGGCC	120
AGTGGCGCGG	CGCGGCGGGG	ACGGCCGCC	AGGCCGCGGT	GGTGCGCTTC	CAAGAACGAG	180
CCAATAAGCA	GAAGCAGGAA	CTCGACGAGA	TCTGACGAA	TATTCGTCAG	GCCGGCGTCC	240
AATACTCGAG	GGCGACGAG	GAGCAGCAGC	AGGCGCTGTC	CTCGCAAATG	GGCTTCTGAC	300
CCGCTAATAC	GAAAAGAAC	GGAGCAA				327

(2) INFORMATION FOR SEQ ID NO:88:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 95 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

Thr	Asp	Ala	Ala	Thr	Leu	Ala	Gln	Glu	Ala	Gly	Asn	Phe	Glu	Arg	Ile
1				5				10					15		
Ser	Gly	Asp	Leu	Lys	Thr	Gln	Ile	Asp	Gln	Val	Glu	Ser	Thr	Ala	Gly

20	25	30
Ser Leu Gln Gly Gln Trp Arg Gly Ala Ala Gly Thr Ala Ala Gln Ala		
35	40	45
Ala Val Val Arg Phe Gln Glu Ala Ala Asn Lys Gln Lys Gln Glu Leu		
50	55	60
Asp Glu Ile Ser Thr Asn Ile Arg Gln Ala Gly Val Gln Tyr Ser Arg		
65	70	75
Ala Asp Glu Glu Gln Gln Ala Leu Ser Ser Gln Met Gly Phe		
85	90	95

TbRa3

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 542 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GAATTCCGCA CGAGAGGTGA TCGACATCAT CGGGACCAGC CCCACATCCT GGGAACAGGC	60
GGCGGGCGAG GCGGTCCAGC GGGCGCGGGA TAGCGTCGAT GACATCCGCG TCGCTCGGGT	120
CATTGAGCAG GACATGGCCG TGGACAGCGC CGGCAAGATC ACCTACCGCA TCAAGCTCGA	180
AGTGTCTTTC AAGATGAGGC CGGCGCAACC GCGCTAGCAC GGGCCGGCGA GCAAGACGCA	240
AAATCGCACG GTTTGCGGTT GATTCTGCG ATTTTGTGTC TGCTCGCCGA GGCCTACCAG	300
GCGCGGCCCA GGTCCGCGTG CTGCCGTATC CAGGCCTGCA TCGCGATTCC GGCGGCCACG	360
CCGGAGTTAA TGCTTCGCGT CGACCCGAAC TGGCGATCC GCCGGNGAGC TGATCGATGA	420
CCGTGGCCAG CCCGTCGATG CCCGAGTTGC CCGAGGAAAC GTGCTGCCAG GCCGGTAGGA	480
AGCGTCCGTA GGCGCGGGTCTGACCGGCT CTGCCCTGCGC CCTCAGTGC GGCAGCGAGC	540
GG	542

(2) INFORMATION FOR SEQ ID NO:77:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 66 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

Val Ile Asp Ile Ile Gly Thr Ser Pro Thr Ser Trp Glu Gln Ala Ala			
1	5	10	15
Ala Glu Ala Val Gln Arg Ala Arg Asp Ser Val Asp Asp Ile Arg Val			
20	25	30	
Ala Arg Val Ile Glu Gln Asp Met Ala Val Asp Ser Ala Gly Lys Ile			
35	40	45	
Thr Tyr Arg Ile Lys Leu Glu Val Ser Phe Lys Met Arg Pro Ala Gln			
50	55	60	
Pro Arg			
65			

38kD

(2) INFORMATION FOR SEQ ID NO:154:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1993 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:154:

TGTTCTTCGA CGGCAGGCTG GTGGAGGAAG GGCCCACCGA ACAGCTGTT	60
AGCATCGGGA AACCAGCCGA TACGTGCCG GACTGTCGGG GGACGTCAG GACGCCAAGC	120
GCGGAAATTG AAGAGCACAG AAAGGTATGG CGTAAAATT CGTTTGCAATA CGCTGTTGGC	180
CGTGTGACCG CCGCGCCGC TGCTGCTAGC AGCGGCGGGC TGTGGCTCGA AACACCGAG	240
CGGTTGCCCT GAAACGGGCG CGGGCGCCGG TACTGTCGCG ACTACCCCG CGTCGTCGCC	300
GGTACGTTG CGGGAGACCG GTAGCACGCT GCTCTACCG CTGTTCAACC TGTGGGGTCC	360
GGCCTTCAC GAGAGGTATC CGAACGTCAC GATCACCGCT CAGGGCACCG GTTCTGGTGC	420
CGGGATCGCG CAGGCCGCCG CGGGGACGGT CAACATTGGG GCCTCCGACG CCTATCTGTC	480
GGAAGGTGAT ATGCCGCGC ACAAGGGCT GATGAACATC GCGCTAGCCA TCTCCGCTCA	540
GCAGGTCAAC TACAACCTGC CGGGAGTGG CGAGCACCTC AAGCTGAACG GAAAAGTCCT	600
GGCGGCCATG TACCAAGGCA CCATCAAAAC CTGGGACGAC CGCGAGATCG CTGGCTCAA	660
CCCCGGCGTG AACCTGCCCCG GCACCGCGGT AGTACCGCTG CACCGCTCCG ACGGGTCGG	720
TGACACCTTC TTGTTCACCC AGTACCTGTC CAAGCAAGAT CCCGAGGGCT GGGGCAAGTC	780
GGCGGCCCTC GGCAACCACCG TCGACTTCCC GCGGTGCCG GGTGCGCTGG GTGAGAACGG	840
CAACGGCGGC ATGGTACCG GTTGCGCCGA GACACGGGC TGCGTGGCCT ATATCGGCAT	900
CAGCTTCTC GACCAGGCCA GTCAACGGGG ACTCGGCGAG GCCCAACTAG GCAATAGCTC	960
TGGCAATTTC TTGTTGCCCG ACGCGCAAAG CATTCAAGGCC GCGGCGGCTG GCTTCGCATC	1020
GAAAACCCCG GCGAACCCAGG CGATTCGAT GATCGACGGG CCCGCCCCGG ACGGCTACCC	1080
GATCATCAAC TACCGAGTACG CCATCGTCAA CAACCGGCCA AAGGACGCCG CCACCGCGCA	1140
GACCTTGCAG GCATTCTGC ACTGGGCGAT CACCGACGGC AACAAAGGCCT CGTTCCCTCGA	1200
CCAGGTTCAT TTCCAGGCCG TGCCGCCCGC GGTGGTGAAG TTGTCTGACG CGTTGATCGC	1260
GACGATTTC AGCTAGCCTC GTTGACCAACC ACCGGACAGC AACCTCCGTC GGGCCATCGG	1320
GCTGCTTGC GGAGCATGCT GGCCCGTGCC GGTGAAGTCG GCCGCGCTGG CCCGGCCATC	1380
CGGTGGTTGG GTGGGATAGG TGCGGTGATC CGCGTCTTG CGCTGGTCTT GGTGCTGGTG	1440
GTGCTGGTCA TCGAGGCCGAT GGGTGCATC AGGCTCAACG GGTTGCATTT CTTCACCGCC	1500
ACCGAATGGA ATCCAGGCCA CACCTACGGC GAAACCGTIG TCACCGACGC GTGCCCATC	1560
CGGTGGCGC CTACTACGGG GCGTTGCCG TGATCGTCGG GACGCTGGCG ACCTCGGCAA	1620
TCGCCCTGAT CATCGCGGT CGCGTCTTG TAGGAGCGGC GCTGGTGATC GTGGAACGGC	1680
TGCCGAAACG GTTGGCCGAG GCTGTGGGAA TAGTCCTGGA ATTGCTCGCC GGAATCCCCA	1740
GCGTGGTCGT CGGTTGTGG GGGGCAATGA CGTCGGGCC GTTCATCGCT CATCACATCG	1800
CTCCGGTGT CGCTCACAAAC GCTCCCGATG TGCCGGTGT GAAACTACTTG CGCGGCGACC	1860
CGGGCAACGG GGAGGGCATG TTGGTGTCCG GTCTGGTGT GGCGGTGATG GTCGTTCCCCA	1920
TTATCGCCAC CACCACTCAT GACCTGTCC GGCAAGGTGCC GGTGTTGCC CGGGAGGGCG	1980
CGATCGGGAA TTC	1993

(2) INFORMATION FOR SEQ ID NO:155:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 374 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:155:

Met Lys Ile Arg Leu His Thr Leu Leu Ala Val Leu Thr Ala Ala Pro
 1 5 10 15
 Leu Leu Leu Ala Ala Ala Gly Cys Gly Ser Lys Pro Pro Ser Gly Ser
 20 25 30
 Pro Glu Thr Gly Ala Gly Ala Gly Thr Val Ala Thr Thr Pro Ala Ser
 35 40 45
 Ser Pro Val Thr Leu Ala Glu Thr Gly Ser Thr Leu Leu Tyr Pro Leu
 50 55 60
 Phe Asn Leu Trp Gly Pro Ala Phe His Glu Arg Tyr Pro Asn Val Thr
 65 70 75 80
 Ile Thr Ala Gln Gly Thr Gly Ser Gly Ala Gly Ile Ala Gln Ala Ala
 85 90 95
 Ala Gly Thr Val Asn Ile Gly Ala Ser Asp Ala Tyr Leu Ser Glu Gly
 100 105 110
 Asp Met Ala Ala His Lys Gly Leu Met Asn Ile Ala Leu Ala Ile Ser
 115 120 125
 Ala Gln Gln Val Asn Tyr Asn Leu Pro Gly Val Ser Glu His Leu Lys
 130 135 140
 Leu Asn Gly Lys Val Leu Ala Ala Met Tyr Gln Gly Thr Ile Lys Thr
 145 150 155 160
 Trp Asp Asp Pro Gln Ile Ala Ala Leu Asn Pro Gly Val Asn Leu Pro
 165 170 175
 Gly Thr Ala Val Val Pro Leu His Arg Ser Asp Gly Ser Gly Asp Thr
 180 185 190
 Phe Leu Phe Thr Gln Tyr Leu Ser Lys Gln Asp Pro Glu Gly Trp Gly
 195 200 205
 Lys Ser Pro Gly Phe Gly Thr Thr Val Asp Phe Pro Ala Val Pro Gly
 210 215 220
 Ala Leu Gly Glu Asn Gly Asn Gly Met Val Thr Gly Cys Ala Glu
 225 230 235 240
 Thr Pro Gly Cys Val Ala Tyr Ile Gly Ile Ser Phe Leu Asp Gln Ala
 245 250 255
 Ser Gln Arg Gly Leu Gly Glu Ala Gln Leu Gly Asn Ser Ser Gly Asn
 260 265 270
 Phe Leu Leu Pro Asp Ala Gln Ser Ile Gln Ala Ala Ala Gly Phe
 275 280 285
 Ala Ser Lys Thr Pro Ala Asn Gln Ala Ile Ser Met Ile Asp Gly Pro
 290 295 300

Ala Pro Asp Gly Tyr Pro Ile Ile Asn Tyr Glu Tyr Ala Ile Val Asn
305 310 315 320

Asn Arg Gln Lys Asp Ala Ala Thr Ala Gln Thr Leu Gln Ala Phe Leu
325 330 335

His Trp Ala Ile Thr Asp Gly Asn Lys Ala Ser Phe Leu Asp Gln Val
340 345 350

His Phe Gln Pro Leu Pro Pro Ala Val Val Lys Leu Ser Asp Ala Leu
355 360 365

Ile Ala Thr Ile Ser Ser
370

DPEP

(2) INFORMATION FOR SEQ ID NO:52:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 999 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

ATGCATCACCA	ATCACCATCA	CATGCATCAG	GTGGACCCCCA	ACTTGACACG	TCGCAAGGGA	60
CGATTGGCGG	CACTGGCTAT	CGCGGCGATG	GCCAGCGCCA	GCCTGGTGAC	CGTTGCGGTG	120
CCC CGC GACCG	CCAAACGCCGA	TCCGGAGCCA	GCGCCCCCGG	TACCCACAAC	GCCCGCCTCG	180
CCGCGCTCGA	CCGCTGCAGC	GCCACCCGCA	CCGGCGACAC	CTGTTGCCCC	CCCACCACCG	240
GCCGCCGCCA	ACACGCCGAA	TGCCCAGCCG	GGCGATCCCA	ACGCAGCACC	TCCGCCGGCC	300
GACCCGAACG	CACCGCCGCC	ACCTGTCATT	GCCCCAAACG	CACCCCAACC	TGTCGGATC	360
GACAACCCGG	TTGGAGGATT	CAGCTTCGCG	CTGCCCTGCTG	GCTGGGTGGA	GTCTGACGCC	420
GCCCCACTTCG	ACTACGGTTC	AGCACTCCTC	AGCAAAACCA	CGGGGGACCC	GCCATTTCAC	480
GGACAGCCGC	CGCCGGTGGC	CAATGACACC	CGTATCGTGC	TCGGCCGGCT	AGACAAAAG	540
CTTTACGCCA	GCGCCGAAGC	CACCGACTCC	AAGGCCGCGG	CCC GGTTGGG	CTCGGACATG	600
GGTGAGTTCT	ATATGCCCTA	CCCGGGCACC	CGGATCAACC	AGGAAACCGT	CTCGCTCGAC	660
GCCAACGGGG	TGTCTGGAAG	CGCGTCGTAT	TACGAAGTCA	AGTCAGCGA	TCCGAGTAAG	720
CCGAACGGCC	AGATCTGGAC	GGCGTAATC	GGCTCGCCCG	CGGCGAACGC	ACCGGACGCC	780
GGGCCCCCTC	AGGCCTGGTT	TGTGGTATGG	CTCGGGACCG	CCAACAAACCC	GGTGGACAAG	840
GGCGCGGCCA	AGGCCTGGC	CGAATCGATC	CGGCCTTTGG	TGCCCCCGCC	GCCGGCGCCG	900
GCACCGGCTC	CTGCAGAGCC	CGCTCCGGCG	CGGGCGCCGG	CGGGGGAAAGT	CGCTCCTACC	960
CCGACGACAC	CGACACCGCA	GCGGACCTTA	CGGGCCTGA			999

(2) INFORMATION FOR SEQ ID NO:53:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 332 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

DSE638572 245268

Met	His	His	His	His	His	Met	Gln	Val	Asp	Pro	Asn	Leu	Thr		
1						5		10				15			
Arg	Arg	Lys	Gly	Arg	Leu	Ala	Ala	Leu	Ala	Ile	Ala	Ala	Met	Ala	Ser
						20		25				30			
Ala	Ser	Leu	Val	Thr	Val	Ala	Val	Pro	Ala	Thr	Ala	Asn	Ala	Asp	Pro
						35		40				45			
Glu	Pro	Ala	Pro	Pro	Val	Pro	Thr	Thr	Ala	Ala	Ser	Pro	Pro	Ser	Thr
						50		55			60				
Ala	Ala	Ala	Pro	Pro	Ala	Pro	Ala	Thr	Pro	Val	Ala	Pro	Pro	Pro	Pro
						65		70			75			80	
Ala	Ala	Ala	Asn	Thr	Pro	Asn	Ala	Gln	Pro	Gly	Asp	Pro	Asn	Ala	Ala
						85		90			95				
Pro	Pro	Pro	Ala	Asp	Pro	Asn	Ala	Pro	Pro	Pro	Val	Ile	Ala	Pro	
						100		105			110				
Asn	Ala	Pro	Gln	Pro	Val	Arg	Ile	Asp	Asn	Pro	Val	Gly	Gly	Phe	Ser
						115		120			125				
Phe	Ala	Leu	Pro	Ala	Gly	Trp	Val	Glu	Ser	Asp	Ala	Ala	His	Phe	Asp
						130		135			140				
Tyr	Gly	Ser	Ala	Leu	Leu	Ser	Lys	Thr	Thr	Gly	Asp	Pro	Pro	Phe	Pro
						145		150			155			160	
Gly	Gln	Pro	Pro	Pro	Val	Ala	Asn	Asp	Thr	Arg	Ile	Val	Leu	Gly	Arg
						165		170			175				
Leu	Asp	Gln	Lys	Leu	Tyr	Ala	Ser	Ala	Glu	Ala	Thr	Asp	Ser	Lys	Ala
						180		185			190				
Ala	Ala	Arg	Leu	Gly	Ser	Asp	Met	Gly	Glu	Phe	Tyr	Met	Pro	Tyr	Pro
						195		200			205				
Gly	Thr	Arg	Ile	Asn	Gln	Glu	Thr	Val	Ser	Leu	Asp	Ala	Asn	Gly	Val
						210		215			220				
Ser	Gly	Ser	Ala	Ser	Tyr	Tyr	Glu	Val	Lys	Phe	Ser	Asp	Pro	Ser	Lys
						225		230			235			240	
Pro	Asn	Gly	Gln	Ile	Trp	Thr	Gly	Val	Ile	Gly	Ser	Pro	Ala	Ala	Asn
						245		250			255				
Ala	Pro	Asp	Ala	Gly	Pro	Pro	Gln	Arg	Trp	Phe	Val	Val	Trp	Leu	Gly
						260		265			270				
Thr	Ala	Asn	Asn	Pro	Val	Asp	Lys	Gly	Ala	Ala	Lys	Ala	Leu	Ala	Glu
						275		280			285				
Ser	Ile	Arg	Pro	Leu	Val	Ala	Pro	Pro	Ala	Pro	Ala	Pro	Ala	Pro	
						290		295			300				
Ala	Glu	Pro	Ala	Pro	Ala	Pro	Ala	Gly	Glu	Val	Ala	Pro	Thr		
						305		310			315			320	
Pro	Thr	Thr	Pro	Thr	Pro	Gln	Arg	Thr	Leu	Pro	Ala				
						325		330							

TbH4

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 702 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

CGGCACGAGG ATCGGTACCC CGCGGCATCG GCAGCTGCCG ATTGGCCGGG TTTCCCCACC

60

CGAGGAAAGC CGCTACCAGA TGGCGCTGCC	GAAGTAGGGC GATCCGTTCG	CGATGCCGGC	120
ATGAACGGGC GGCACTAAAT TAGTGCAGGA ACCTTCAGT TTAGCGACGA	TAATGGCTAT		180
AGCACTAAGG AGGATGATCC GATATGACGC AGTCGCAGAC	CGTGACCGTG GATCAGCAAG		240
AGATTGGAA CAGGGCCAAC GAGGTGGAGG CCCCGATGGC	GGACCCACCG ACTGATGTCC		300
CCATCACACC GTGCGAACTC ACGGNGNTA AAAACGCCGC	CCAACAGNTG GTNTTGTCCG		360
CCGACAAACAT GCGGAAATAC CTGGCGCCCG GTGCCAAAGA	GCGGCAGCGT CTGGCGACCT		420
CGCTCGCAA CGCGGCCAAAG GNNTATGGCG AGGTTGATGA	GGAGGCTGCG ACCCGCTGG		480
ACAACGACGG CGAAGGAACT GTGCAGGCAG AATCGGCCGG	GGCGTCCGA GGGGACAGTT		540
CGGCCGAACT AACCGATAAG CCGAGGGTGG CCACGGCCGG	TGAACCCAAC TTCATGGATC		600
TCAAAGAACG GCGAAGGAAG CTCGAAACGG GCGACCAAGG	CGCATCGCTC GCGCACTGNG		660
GGGATGGGTG GAACACTTNC ACCCTGACGC TGCAAGGCAG	CG		702

(2) INFORMATION FOR SEQ ID NO:81:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 286 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

```

Gly Asp Ser Phe Trp Ala Ala Ala Asp Gln Met Ala Arg Gly Phe Val
1      5          10        15
Leu Gly Ala Thr Ala Gly Arg Thr Thr Leu Thr Gly Glu Gly Leu Gln
20     25          30
His Ala Asp Gly His Ser Leu Leu Asp Ala Thr Asn Pro Ala Val
35     40          45
Val Ala Tyr Asp Pro Ala Phe Ala Tyr Glu Ile Gly Tyr Ile Xaa Glu
50     55          60
Ser Gly Leu Ala Arg Met Cys Gly Glu Asn Pro Glu Asn Ile Phe Phe
65     70          75        80
Tyr Ile Thr Val Tyr Asn Glu Pro Tyr Val Gin Pro Pro Glu Pro Glu
85     90          95
Asn Phe Asp Pro Glu Gly Val Leu Gly Gly Ile Tyr Arg Tyr His Ala
100    105         110
Ala Thr Glu Gln Arg Thr Asn Lys Xaa Gln Ile Leu Ala Ser Gly Val
115    120         125
Ala Met Pro Ala Ala Leu Arg Ala Ala Gln Met Leu Ala Ala Glu Trp
130    135         140
Asp Val Ala Ala Asp Val Trp Ser Val Thr Ser Trp Gly Glu Leu Asn
145    150         155        160
Arg Asp Gly Val Val Ile Glu Thr Glu Lys Leu Arg His Pro Asp Arg
165    170         175
Pro Ala Gly Val Pro Tyr Val Thr Arg Ala Leu Glu Asn Ala Arg Gly
180    185         190
Pro Val Ile Ala Val Ser Asp Trp Met Arg Ala Val Pro Glu Gln Ile
195    200         205
Arg Pro Trp Val Pro Gly Thr Tyr Leu Thr Leu Gly Thr Asp Gly Phe
210    215         220
Gly Phe Ser Asp Thr Arg Pro Ala Gly Arg Arg Tyr Phe Asn Thr Asp
225    230         235        240
Ala Glu Ser Gln Val Gly Arg Gly Phe Gly Arg Gly Trp Pro Gly Arg
245    250         255
Arg Val Asn Ile Asp Pro Phe Gly Ala Gly Arg Gly Pro Pro Ala Gln
260    265         270
Leu Pro Gly Phe Asp Glu Gly Gly Leu Arg Pro Xaa Lys

```

DPPD DPPD DPPD DPPD

275

280

285

MTbRa12

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 447 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CGGTATGAAC	ACGGCCGCGT	CCGATAACTT	CCAGCTGTCC	CAGGGTGGGC	AGGGATTTCGC	60
CATTCCGATC	GGGCAGGCAG	TGGCGATCGC	GGGCCAGATC	CGATCGGGTG	GGGGGTCACC	120
CACCGTTCAT	ATCGGGCCTA	CCGCCTTCCT	CGGCTTGGGT	GTTGTGGACA	ACAAACGGCAA	180
CGGCGCACGA	GTCCAACGCG	TGGTCGGGAG	CGCTCCGGCG	GCAAGTCTCG	GCATCTCCAC	240
CGGCGACGTG	ATCACCGCGG	TCGACGGCGC	TCCGATCAAC	TCGGCCACCG	CGATGGCGGA	300
CGCGCTTAAC	GGGCATCATC	CCGGTGACGT	CATCTCGGTG	AACTGGAAA	CCAAGTCGGG	360
CGGCACCGCGT	ACAGGGAACG	TGACATTGGC	CGAGGGACCC	CCGGCCTGAT	TTCGTCGYGG	420
ATACCACCCG	CCGGCCGGCC	AATTGGA				447

(2) INFORMATION FOR SEQ ID NO:66:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 132 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

Thr	Ala	Ala	Ser	Asp	Asn	Phe	Gln	Leu	Ser	Gln	Gly	Gly	Gln	Gly	Phe
1						5				10				15	
Ala	Ile	Pro	Ile	Gly	Gln	Ala	Met	Ala	Ile	Ala	Gly	Gln	Ile	Arg	Ser
									20		25		30		
Gly	Gly	Gly	Ser	Pro	Thr	Val	His	Ile	Gly	Pro	Thr	Ala	Phe	Leu	Gly
									35		40		45		
Leu	Gly	Val	Val	Asp	Asn	Asn	Gly	Asn	Gly	Ala	Arg	Val	Gln	Arg	Val
									50		55		60		
Val	Gly	Ser	Ala	Pro	Ala	Ala	Ser	Leu	Gly	Ile	Ser	Thr	Gly	Asp	Val
									65		70		75		80
Ile	Thr	Ala	Val	Asp	Gly	Ala	Pro	Ile	Asn	Ser	Ala	Thr	Ala	Met	Ala
									85		90		95		
Asp	Ala	Leu	Asn	Gly	His	His	Pro	Gly	Asp	Val	Ile	Ser	Val	Asn	Trp
									100		105		110		
Gln	Thr	Lys	Ser	Gly	Gly	Thr	Arg	Thr	Gly	Asn	Val	Thr	Leu	Ala	Glu
									115		120		125		
Gly	Pro	Pro	Ala												
			130												

DPPD

(2) INFORMATION FOR SEQ ID NO:240:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 339 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:240:

ATGAAGTTGA AGTTTGCTCG CCTGAGTACT GCGATACTGG GTTGTGCAGC GGCGCTTGTG	60
TTTCCTGCCT CGGTTGCCAG CGCAGATCCA CCTGACCCGC ATCAGCCGGA CATGACGAAA	120
GGCTATTGCC CGGGTGGCCG ATGGGGTTTT GGCGACTTGG CCGTGTGCGA CGGCGAGAAG	180
TACCCCGACG GCTCGTTTG GCACCAGTGG ATGCAAACGT GGTTTACCGG CCCACAGTTT	240
TACTTCGATT GTGTCAGCGG CGGTGAGCCC CTCCCCGGCC CGCCGCCACC GGGTGGTTGC	300
GGTGGGGCAA TTCCGTCCGA GCAGCCAAAC GCTCCCTGA	339

(2) INFORMATION FOR SEQ ID NO:241:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 112 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:241:

Met Lys Leu Lys Phe Ala Arg Leu Ser Thr Ala Ile Leu Gly Cys Ala			
1	5	10	15
Ala Ala Leu Val Phe Pro Ala Ser Val Ala Ser Ala Asp Pro Pro Asp			
20	25	30	
Pro His Gln Pro Asp Met Thr Lys Gly Tyr Cys Pro Gly Gly Arg Trp			
35	40	45	
Gly Phe Gly Asp Leu Ala Val Cys Asp Gly Glu Lys Tyr Pro Asp Gly			
50	55	60	
Ser Phe Trp His Gln Trp Met Gln Thr Trp Phe Thr Gly Pro Gln Phe			
65	70	75	80
Tyr Phe Asp Cys Val Ser Gly Gly Glu Pro Leu Pro Gly Pro Pro Pro			
85	90	95	
Pro Gly Gly Cys Gly Gly Ala Ile Pro Ser Glu Gln Pro Asn Ala Pro			
100	105	110	

ESAT-6

(2) INFORMATION FOR SEQ ID NO:103:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 154 base pairs
- (B) TYPE: nucleic acid

- (C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:103:

ATGACAGAGC AGCAGTGGAA TTTCGCGGGT ATCGAGGCCG CGGCAAGCGC AATCCAGGGA 60
AATGTACGT CCATTCAATTC CCTCCTTGAC GAGGGGAAGC AGTCCCTGAC CAAGCTCGCA 120
GCGGCCTGGG CGGGTAGCGG TTCTGGAAGCG TACC 154

(2) INFORMATION FOR SEQ ID NO:104:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 51 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:104:

Met Thr Glu Gln Gln Trp Asn Phe Ala Gly Ile Glu Ala Ala Ala Ser
1 15
 15
Ala Ile Gln Gly Asn Val Thr Ser Ile His Ser Leu Leu Asp Glu Gly
20 30
 30
Lys Gln Ser Leu Thr Lys Leu Ala Ala Ala Trp Gly Gly Ser Gly Ser
35 45
 45
Glu Ala Tyr
50